A Bayesian Analysis of Landscape Disturbance and the Ecosystem Enhancement of Q'eqchi' Maya Swidden Agriculture

Presented in partial fulfillment of the requirements for graduation with honors research distinction in the undergraduate colleges of The Ohio State University

By

Matthew Walker

Undergraduate

The Ohio State University

April 2021

Project Advisor: Dr. Sean S. Downey, Department of Anthropology

Committee members: Dr. Mark Moritz, Dr. William Peterman

Table of Contents

Abstract

In this paper we ask whether forest clearing patterns and the regrowth associated with swidden cultivation exhibits patterns of disturbance that enhance ecosystem productivity. Analyzing high-resolution, multispectral imagery and ground-truthed data collected in 2018 from two villages in southern Belize, we test the Intermediate Disturbance Hypothesis (IDH) by relating spatial metrics of landscape fragmentation to diversity and abundance of tree species. The analysis workflow outlined in this paper makes use of high-performance computing (HPC) to overcome constraints of scale, data size, and algorithmic complexity. This effectively cuts the time to process data and produce statistical results by two orders of magnitude, enabling us to use more complex algorithms, to analyze larger high-precision datasets, and to rapidly iterate our analyses. Preliminary modeling results produced by this analysis qualitatively reveal significant differentiation between anthropogenic disturbances expanding from the village centers and the primary forest and provide tentative quantitative support for the claim that Q'eqchi' forest clearing and swidden cultivation enhances forest biodiversity in southern Belize, especially at large spatial scales of analysis.

1. Introduction

Studies of conservation in subsistence-oriented, small scale, and indigenous societies in tropical parts of the world often focus on ascertaining whether swidden agriculture – also known as "slash-and-burn" – is inherently destructive of tropical forests or whether it can exhibit sustainability. We can define sustainability here as the use and consumption of forest resources over time without loss to species diversity and richness. To determine whether these agricultural strategies can generate long-term sustainability or merely short-term benefits, we examine the effects of swidden agriculture practices on landscape fragmentation by analyzing very high resolution multispectral remote sensing data, spatial statistics, and ground-truthed data to detect quantitative signatures that may link the traditional ecological knowledge (TEK) of the Q'eqchi Maya in southern Belize with local ecological sustainability. Reaching an answer to this question necessitates an analytical, high-performance computing approach due to the large size of the data and the complexity of models being evaluated.

As a part of ongoing research with the Human Complexity Lab at The Ohio State University, a high-resolution multispectral scan was captured in 2018 by a drone equipped with a MicaSense RedEdge sensor over the area encompassing the villages Crique Sarco and Graham Creek in southern Belize. The sensor data includes five bands: Blue, Green, Red, Red Edge, and NIR.

Team members used unsupervised learning to stitch together over 36,000 individual photographs into a multi-band image which contains spectral signatures. The images were validated and classified using ground-truthed GPS survey data from farm plots, ranging from primary forest (>50 years since disturbance) to <1 year since disturbance.

Using the spectral signatures as a proxy for local species abundance (including α-diversity, β-diversity), my role on the project was to create an analysis workflow that efficiently samples the relevant statistics and links this analysis with that of the land-use classification. I then modeled these results using a Bayesian approach that relates metrics of seral stage diversity and environment heterogeneity at varying levels (patch, class, landscape) to species abundance, showing how these landscapes are affected by human activity and how forests respond to different patterns and extent of swidden disturbance patterns.

Since January 2020, as an undergraduate research assistant, I have been responsible for developing computationally-intensive statistical methods to measure the spatial fragmentation and biodiversity of mosaic swidden systems in the tropical rainforests of southern Belize, with a focus on improving the operational time to analysis via novel approaches to big data problems, which is helping the team to create workflows that push the limits of what was previously possible. I pursued a B.S. in Data Analytics with a specialization in Social Science Analytics (SSA), gaining experience both in academic and industry settings. I am interested in growing interdisciplinary partnerships across the social sciences and bringing statistical methodologies to new domains. Over the course of 2020, I contributed mostly to the implementation and testing of a comprehensive adaptation of class- and landscape-level fragmentation analyses on a classified raster. After visualizing the results of these spatial statistics, I successfully parallelized the workflow in a distributed computing environment, cutting the time from classification to insights about system fragmentation by over two orders of magnitude. Using high-resolution remote sensing imagery captured from a drone in the field, I developed an approach to analyzing the

problem of landscape diversity from a different angle – unsupervised learning – identifying "species" by their spectral signatures. Connecting the two analyses will pave the way for future application of high-performance computing and big data methodologies to anthropological questions.

The overarching research question explored in this thesis is whether computational methods can be used to detect a quantitative signature of ecosystem enhancement at intermediate levels of swidden disturbance. If so, this suggests that subsistence-oriented swidden cultivators are acting as trophic mediators in ecosystems characterized by human-environmental interactions, mediated by social norms, and constrained by ecological dynamics.

2. Background

The potential merits – and risks – of swidden agriculture, where farmers cultivate subsistence crops such as maize in patches of forests cleared by a process commonly known as "slash-and-burn," have been widely debated across disciplines. Often, these discussions are focused on ascertaining whether swidden is inherently destructive of tropical forest environments (Malthus 1826, Boserup 1965), or whether it can exhibit sustainability (Downey et al. 2020; Geertz 1963; Downey 2010). The focus of this research project is to analyze whether spatial patterns of human disturbance caused by swidden agricultural practices in tropical forest canopies contain a quantitative signature of ecological resilience; that is, at an intermediate spatial level, can anthropogenic disturbances enhance landscape structure, function, and productivity?

Two predominant, opposing theories have emerged as potential explanations for the complex functions of swidden agriculture. The first argues that swidden is strictly limited by ecological carrying capacity, or the maximum sustainable population size given the availability of resources, and therefore subjects natural environments to overexploitation and deforestation (Food and Agriculture Organization 1985). This approach is often used in discourse about how to best manage forest landscapes. The second approach, contrary to 'Western scientific' management practices, conceives of 'coupled human and natural systems' in which human cultures develop detailed knowledge of their local environments and the processes which best

conserve and protect this ecosystem – a collective understanding referred to as Traditional Ecological Knowledge, or TEK (Berkes 2017). However, despite the significant corpora of research, which span questions about the integration of swidden in ecology, culture, society, and religion (Conklin 1954, Geertz 1963), as well as developing sustainable swidden and augmenting forest resilience (Balée 2013) and the larger role swidden systems play in a globalized, technological world (Weinstock 2015), the relationships between specific social and spatial dynamics and ecological features such as carrying capacity remain unclear. However, despite the negative connotation of "slash-and-burn," we do know that swidden agriculture is how many indigenous communities like the Q'eqchi Maya of southern Belize have been sustainably producing food for thousands of years (Downey 2010, 2020).

There are a number of complex cultural and historical contexts which are needed to analyze the swidden practices of Q'eqchi' Maya communities. Importantly, Q'eqchi' social institutions do not formally protect common forest resources (Downey 2009). The resources are instead collectively managed through reciprocal exchanges of labor, among other cultural practices and social norms. During the colonial period, these practices emerged when access to government land for swidden production – both subsistence and market-oriented – was largely unregulated, making land a common property resource available to Q'eqchi' farmers (Downey 2015). In most villages, labor exchanges – which occur when one farmer asks a group of men to help with a difficult task such as planting or clearing the forest – have become the key social institution involved with swidden agriculture (Wilk 1997). After each shared job, or workgroup, a farmer is expected to close his debt to each man who helped him by reciprocating a day of labor. This reciprocity effect and the communal nature of agricultural practices in the Q'eqchi' villages has already been observed and reproduced using experimental games (Downey et al. 2020) and work logs (Downey, unpublished data). But linking these individual actors and actions to a larger framework of socialization and environmental impact demands a cross-scale analysis; one that can measure complex patterns of landscape disturbance and relate these metrics back to the practices of small-scale subsistence farmers in Crique Sarco and Graham Creek.

The long-term goal of this project is to link social and cultural norms with tropical forest ecology. One lens for exploring this complex relationship is resilience, or how well the landscape ecology can resist and recover from disturbances at different scales. Although we lack general

analytical tools to evaluate variation in resilience, particularly – sometimes by design – in social portions of coupled systems, we can still benefit from analyzing real systems by applying a more quantitative approach.

What distinguishes the methods outlined in this paper from past research in swidden is the need for high-performance computation. The combination of extremely high spatial resolution (0.28 m^2 per pixel), the spatial scale of the raster images (approximately 18,000 acres), the algorithmic complexity of the landscape fragmentation calculations, spectral species identification, and Bayesian modeling together necessitates high performance computing. . In combination with the collaborative, ground-truthed data collection, the project presents a key linkage between two levels of analysis typically done independently of one another. Swidden has rarely been analyzed at this pair of scales: a scale small enough to detect individual (spectral) species, and also large enough to characterize the landscape at the scale that is relevant to human swidden activities (Levin 1992).

To investigate the impacts of landscape fragmentation as a result of small-scale anthropogenic disturbances on species abundance metrics, we investigate whether the intermediate disturbance hypothesis (IDH), predicts the distribution of tropical forest species diversity in fragmented swidden mosaics. The IDH posits that local species diversity is maximized when ecological disturbances are neither too rare nor too frequent and of moderate spatial scale (Connell 1978). Importantly, while disturbances of large extent risk exterminating most species and small disturbances fail to provide sufficient niche partitioning, intermediate disturbances help to facilitate coexistence between colonizer species – which thrive at early successional stages – and other species that out-compete colonizer species during later successional stages (Wilkinson 1999). At the conceptual level, the IDH lays a foundation for beginning to understand broader questions about whether modulating forest fragmentation via swidden disturbances can enhance resource availability.

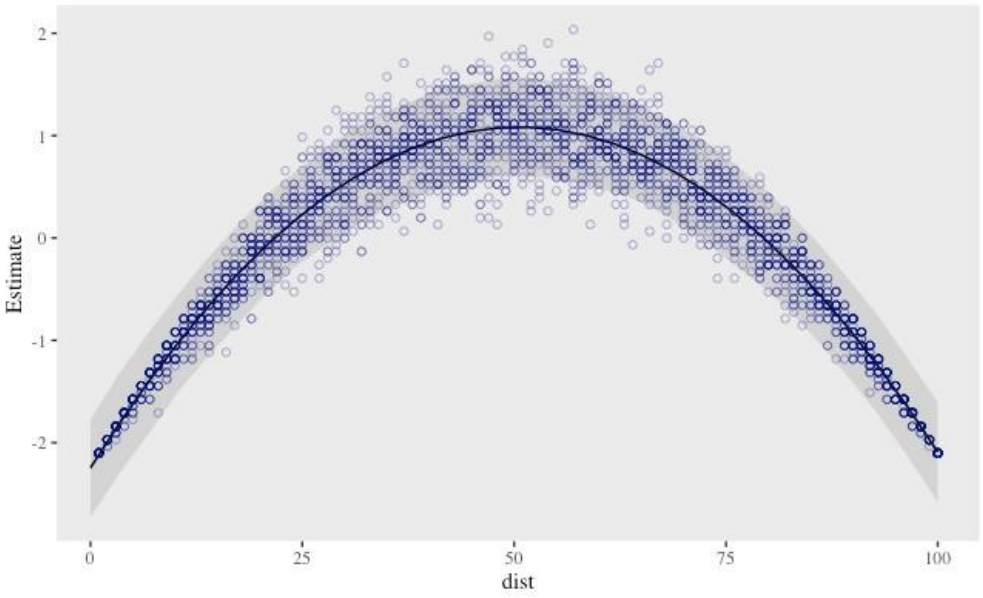


Figure 1: A simulated example of the second-order dynamics predicted by the IDH in which resource availability is maximized at intermediate disturbance intensity with Bayesian posterior predictions (black) and a 95% credibility interval (grey)

While evidence exists both for and against the IDH (Fox 2013), there are numerous findings in support of the IDH in tropical ecosystems. For example, a study of frugivorous birds in the Bolivian Andes determined that species abundance and richness were maximized at forest edges and dispersed seeds were less clustered at edges than in the interior (Saavedra et al. 2014). There are numerous disturbance processes, natural and anthropogenic, which can potentially increase the number of edges in a landscape and thus enhance the productivity of local species. In Australia, disturbance patterns created by Aboriginal burning were found to have optimal effects on hill kangaroo populations at intermediate levels (Codding 2014). Other species that have been studied under the IDH include sand monitor lizards (Bird 2008) and Northwest clam gardens (Cox 2019). These findings support the hypothesis that, over long periods of time, anthropogenic disturbances allow humans to act as trophic mediators in coupled human and natural systems.

In a global context, improving existing models of subsistence agriculture could have significant impacts on policy directed at people and working forests, which comprise nearly two-thirds of all forests worldwide (FAO 2010). Ultimately, findings from this research could illuminate the

complex array of factors which promote effective management of natural resources at the local level. A cohesive understanding of these processes could lay the foundation for healthier tropical forests worldwide while protecting indigenous knowledge.

2.1 Transdisciplinary research

A range of quantitative and qualitative data were available to develop this analysis when I joined the project, including household surveys, farmer land use interviews, and high-resolution, multi-spectral drone data that the team at HCL had stitched together at earlier stages in the project. The data curation process was often in motion, involving expertise from individuals across numerous disciplines and departments. Throughout the duration of my research, the data have gone through many iterations, typically weeks apart due to the time it takes to identify areas of improvement, whether this means manually checking GPS coordinates or reviewing confusion matrices of a new classification. Any code had to be adaptable to new inputs and methodologies.

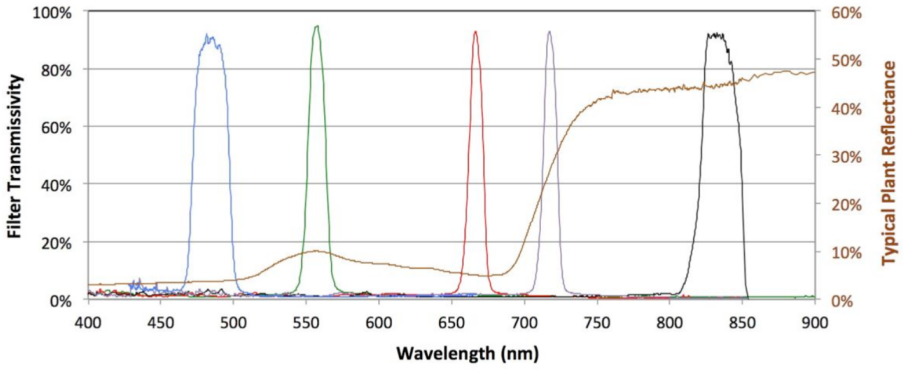
Conducting this type of research requires teams of scientists and community members to work together to address complex problems of societal importance. The accumulation and storage of data for this project was designed to be a collective process involving the input and assistance of local Q'eqchi' experts in the study area as well as cross-disciplinary, collaborative efforts in constructing the analysis.

This was a transdisciplinary effort that could not be completed without assistance from local fieldworkers, GPS mappers, and drone specialists. The photos, once stitched together into geo-referenced, three-dimensional photo-mosaics and verified by farm plots surveys, can be classified and analyzed to understand how these environments provide for the local communities, how they are affected by human activity, and how forests persist over time.

2.2 Drone-based data collection

In April 2018, two FireFLY6 drones and a Micasense RedEdge sensor (see Fig. 2) were used to collect over 35,000 multispectral images encompassing a region of approximately 18,000 acres (~7,200 hectares). In order to achieve the desired spatial coverage, the drone flew at approximately 1,500 feet over a period of two weeks during the dry season when weather was

least likely to interfere with flight operations (Moschler 2018). A research permit was obtained from the Belize Department of Civil Aviation and flight protocols were used to ensure safety of field assistants, local residents, and local air traffic.



Band Number	Band Name	Center Wavelength (nm)	Bandwidth FWHM (nm)
1	Blue	475	20
2	Green	560	20
3	Red	668	10
4	Near IR	840	40
5	Red Edge	717	10

Figure 2: Spectral bands captured by the MicaSense RedEdge sensor and their specifications (Micasense 2015)

2.3 Ground-truthed data

Simultaneous with the drone flight operations in 2018, the PI conducted ground-truthed mapping with a team of surveyors from the study villages who were also farmers and therefore knowledgeable about local land use history. Surveyors were sent out singly or in pairs with GPS devices and instructed to collect coordinate and land-use information including an age-estimate, land use class, and any historical information about the current or past farmers who have used the land.

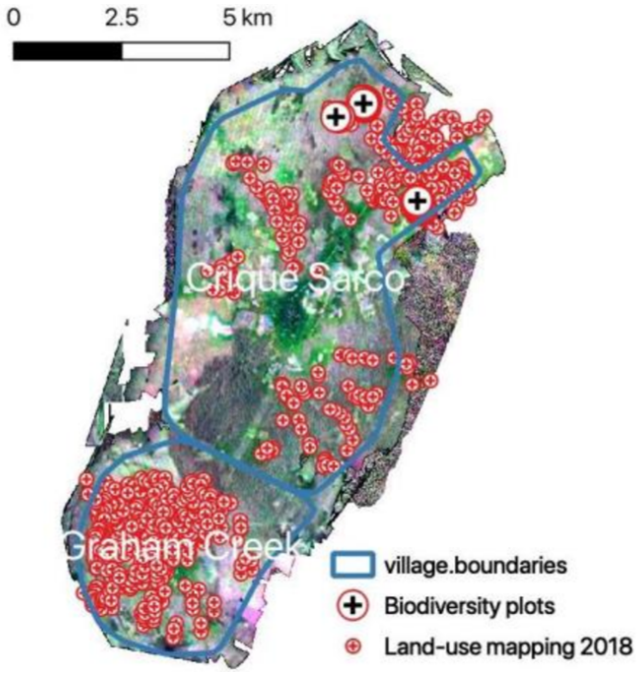


Figure 3: Drone survey data and land-use mapping (Toledo District, Belize)

Surveyors were asked to pay special attention to mapping abandoned fields (known locally as *huamil*) from a wide range of age classes and spatial locations, as well as primary forest areas at the perimeter of the village use areas . These data were manually checked for consistency resulting in a training dataset containing 740 land-use samples. This dataset was then used for training the random forest model (Breiman 2001). The photogrammetry and classification workflows, whose design and implementation were spearheaded by Dr. Rongjun Qin and PhD student Shuang Song (The Ohio State University, College of Engineering), are illustrated in Fig. 4 but the full methodology is beyond the scope of this paper, where we will mainly be focusing on analysis stemming from the final output of these workflows: the classified map.

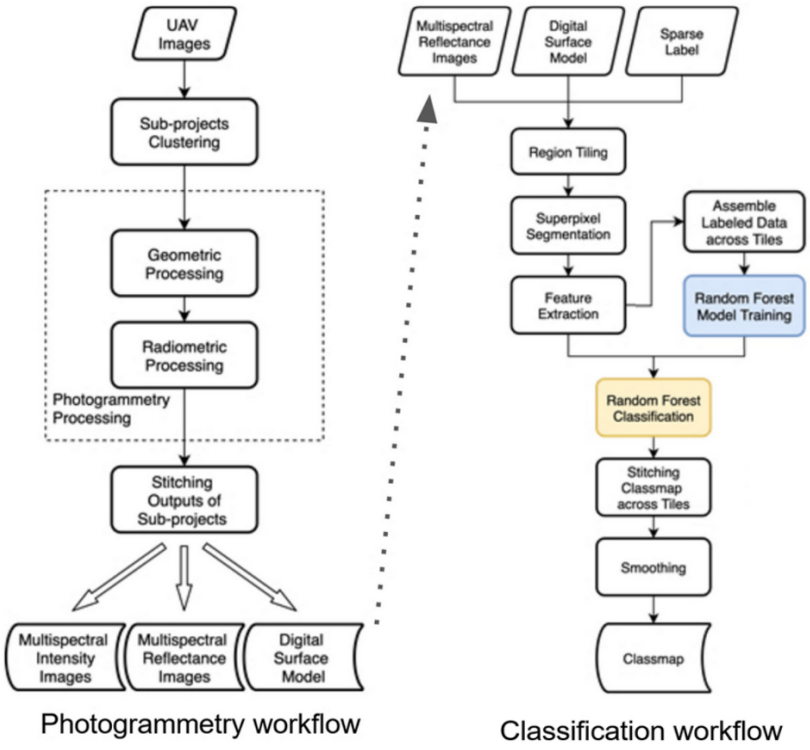


Figure 4: The complete photogrammetry and classification workflows from raw UAV imagery data to the classified landscape raster

Lastly, the PI, PhD student Shane Scaggs, and Juan Pop (Crique Sarco, Community Ecologist) conducted biodiversity assessments at four locations where species abundance was counted by hand to create accurate biodiversity reference points that can be used to validate results (see Appendix 1).

3. Approach

Our analysis approaches testing the hypothesis is outlined in Fig. 5. First, given a classified land-use map detailing the age of various patches (groups of homogeneous cells), we use the R/landscapemetrics package to measure landscape fragmentation and connectivity via statistics at the class and landscape levels, where connectivity refers to the degree to which a landscape facilitates or impedes movement between resource patches (Taylor 1993). A second, parallel approach uses the multispectral reflectance images and R/biodivMapR package to identify spectral species abundance and diversity. Both of the outputs share one polygon sampling grid

for which metrics are generated, making the scales of analysis comparable across both points of approach. Finally, the IDH is tested by relating landscape productivity to fragmentation, with the goal of eventually incorporating above ground biomass (calculated from the digital surface model) as well as additional social and geographic variables.The data needed to accomplish this include:

- 2018 multispectral remote sensing data (RGB, RedEdge, NIR)
- 2018 ground-truthed GPS mapping data (land-use classes)
- Sampling framework: Polygon file for Crique Sarco, Graham Creek, Primary Forest region boundaries
- 2018 transect biodiversity counts in 4 locations

3.1 High-Performance Computing (HPC) workflow

The workflow I developed analyzes two of the spatial datasets: the first is a fragmentation analysis of the automated classification of land-use classes; and the second is a principal components analysis of the spectral variation in multispectral raster images to infer forest diversity from remote sensing data. Due to the size and scale of the data, we took a parallel approach to computing the relevant statistics, which significantly reduced the time to generate new analysis or iterate over existing outputs. The Ohio Supercomputer Center was instrumental in making this workflow possible.

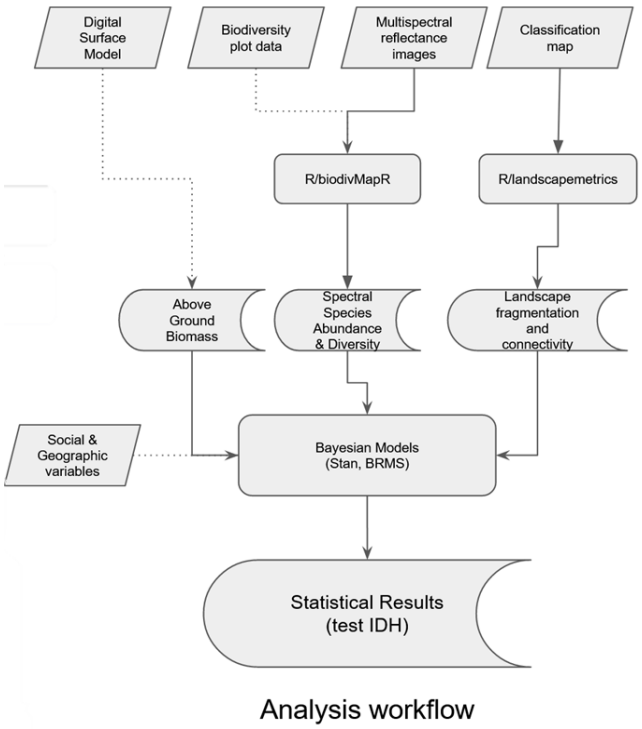


Figure 5: The complete analysis workflow diagram

Producing the final statistics and plots was both computationally expensive and time-consuming, sometimes taking upwards of 20 or more hours to complete on the full raster. To combat this, I developed and tested code on small subsets of the full raster until it was working as expected. Parallelizing the process of calculating the spatial statistics in R also helped significantly, reducing overall computation time to under 1 hour in most cases. High Performance Computing (HPC) was instrumental in making this a possibility, as the data, once split into smaller cells, could be distributed across many cores without significant loss of information when re-combined. Successfully implementing and iterating upon this workflow was a technical challenge, but one that has dramatically improved the time it takes to analyze future data

There are two main components to the landscape-level spatial analysis: fragmentation and biodiversity.

3.2 Fragmentation analysis

In this analysis we consider how farmers' land-use choices result in a landscape that may enhance certain ecological dynamics like seed dispersal (Saavedra et al. 2014). Human activity can optimize variables like seed dispersal by where they position/place their fields, which affects the degree to which the landscape becomes fragmented over time (Jaeger 2000) and the degree to which the resulting landscapes maintain connectivity (Tischendorf 2000, Kindlmann 2008). We want to test whether there is a spatial scale -- or scales -- that encourages, maintains, or enhances seed dispersal and animal activity. Here we can use statistics generated with the landscape metrics package in R to address this question (Hesselbarth 2019).

Using the classified land-use map, we generated two types of landscape statistics: landscape-scale statistical characterization including Shannon's diversity index, patch area, and edge density; and class-level characterizations including patch density, shape index, and radius of gyration.

Class level metrics summarize all patches belonging to a particular class within a focal landscape. These metrics can be either the "distribution" of patch level metrics of all patches of a class (e.g. the mean) or consider only patches of a particular class for the calculations of the metric, making them suitable to describe the composition and configuration of the landscape (Hesselbarth 2019).

Landscape level metrics summarize the whole landscape in one value (in our case, the "landscape" is a whole sampling polygon). This is done by summarising metrics of lower levels or calculating a metric including all patches and classes. Landscape level metrics are suitable for condensing information at a larger scale (Hesselbarth 2019).

The metrics chosen were selected for both ecological relevance and to account for as much variation in the landscape as possible without redundancy (Cushman 2008). Cushman et. al (2008) suggest that there exists a minimum set of structure attributes which can adequately describe the structure patterns within any landscape, and that quantifying more than one or few metrics within each component group is highly redundant. At the class level, seven components of the PCA were nearly universal and consistent in interpreted meaning, while at the landscape

level, eight components were identified as universal and consistent (Cushman 2008). While some metrics, such as total edge contrast index, have not yet been implemented in the landscape metrics package (Hesselbarth 2019), we identified a sufficient set of metrics from each of the most important components where possible. All metrics used and their respective specifications can be found in Table 1 and Table 2.

Metric	Name	Type	Scope
gyrate_mn	Mean radius of gyration	Area and Edge	Class
lsi	Landscape shape index	Aggregation	Class
pd	Patch density	Aggregation	Class
area_mn	Mean patch size	Area and Edge	Class
area_cv	Patch size coefficient of variation	Area and Edge	Landscape
split	Splitting index	Splitting	Landscape
area_mn	Mean patch size	Area and Edge	Landscape
ed	Edge density	Area and Edge	Landscape
shdi	Shannon's diversity index	Diversity	Landscape

Legend	
Frag. Analysis	
Spec. Analysis	
Both Analyses	

Table 1: chosen landscape metrics colored by use, where yellow denotes use in the fragmentation analysis, blue denotes use in the spectral biodiversity analysis, and green denotes both.

Metric	Calculation	Units	Range	Behavior
Mean radius of gyration (c)	$GYRATE_{MN} = mean(GYRATE[patch_{ij}])$	Meters	GYRATE_MN>= 0	Approaches GYRATE_MN=0 if every patch is a single cell. Increases, without limit, when only one patch is present
Landscape shape index (c)	$LSI = \frac{e_i}{min\ e_i}$ where e_i is the total edge length in cell surfaces and $min\ e_i$ is the minimum total edge length in cell surfaces	None	LSI >= 1	Equals LSI=1 when only one squared patch is present or all patches are maximally aggregated. Increases, without limit, as length of actual edges increases, i.e. patches become less compact.

Patch density (c)	$PD = \frac{n_i}{A} * 10000 * 100$ where n_i is the number of patches and A is the total landscape area in square meters	Number per 100 hectares	0 < PD <= 1e+06	Increases as the landscape gets more patchy. Reaches its maximum if every cell is a different patch.
Mean patch size (c) Mean patch size (l)	$AREA_{MN} = mean(AREA[patch_{ij}])$ where $AREA[patch_{ij}]$is the area of each patch in hectares	Hectares	AREA_MN > 0	Approaches AREA_MN=0 if all patches are small. Increases, without limit, as the patch areas increase.
Patch size coefficient of variation (l)	$AREA_{CV} = cv(AREA[patch_{ij}])$ where the *coefficient of variation* $cv = \frac{SD}{MN} * 100$	Hectares	AREA_CV >= 0	Equals AREA_CV=0 if all patches are identical in size. Increases, without limit, as the variation of patch areas increases.
Splitting index (l)	$SSPLIT = \frac{A^2}{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$	None	1 <= SPLIT <= Number of cells squared	Equals SPLIT=1 if only one patch is present. Increases as the number of patches increases and is limited if all cells are a patch.
Edge density (l)	$ED = \frac{E}{A} * 10000$ where E is the total landscape edge in meters and A is the total landscape area in square meters	Meters per hectare	ED >= 0	Equals ED = 0 if only one patch is present (and the landscape boundary is not included) and increases, without limit, as the landscape becomes more patchy.
Shannon's diversity index (l)	$SHDI = -\sum_{i=1}^{m}(P_i * ln\, P_i)$ where P_i is the proportion of class i	None	SHDI >= 0	Equals SHDI=0 when only one patch is present and increases, without limit, as the number of classes increases while the proportions are equally distributed.

Table 2: Chosen landscape metrics as described in (McGarigal 1995)

We calculated these statistics for both villages and for a large region of primary forest that was within the drone scanning area (Fig. 3).

Shannon diversity index, which quantifies the uncertainty in predicting species abundance, can be defined as:

$$H' = -\sum_{i=1}^{S} p_i ln(p_i)$$

where p_i is the proportion of individuals belonging to the ith species in group S. The "species" in this case are the various land-classes of our map: for example, a patch labeled as 1-4 years old. This age represents how much time has elapsed since the last clearing of a particular area, as of 2018 when the data were collected.

To estimate variation in Shannon diversity across the classified raster, the raster was divided into a sampling grid with a window size (see section 3.5, pp. 28), set in the function *sample_lsm* (Hesselbarth 2019), and the desired statistics were then calculated in parallel on each cell (or a subset of cells, depending on available CPUs and RAM.

One of the early iterations of a classification map resulted in an array of calculated values that could be plotted with respect to the sampling cell size and position to visualize variation across regions.

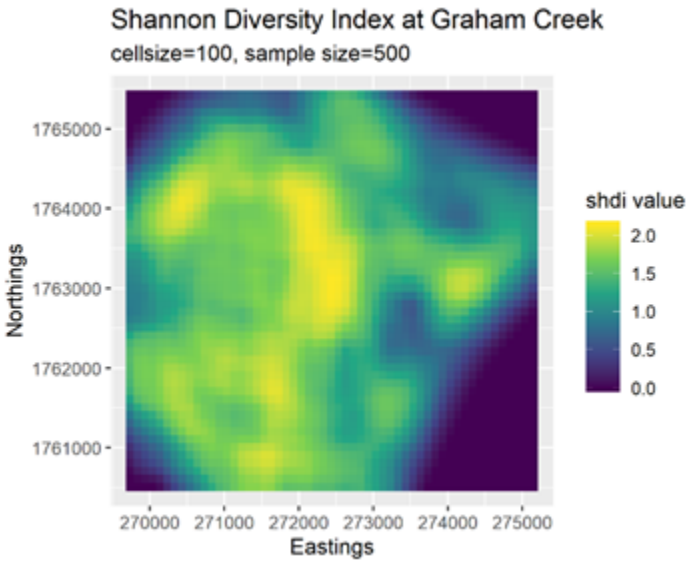


Figure 6: Calculated SHDI values across early classification of Graham Creek region

Other metrics, such as edge density and patch size, can provide additional information about the structure of the landscape patterns that is relevant to the ecological interpretation. For example,

seed dispersal is often found to be optimal in patchy, diverse mosaic systems which exhibit high edge density and low mean patch size.

This fragmentation analysis will not and cannot ‘prove’ ecosystem enhancement (which will require [spectral] species counts), so the goal is to characterize the fragmentation patterns in anthropogenic and non-anthropogenic parts of the forest. These analyses will tie back into the IDH and identify a plausible 'intermediate' scale of disturbance and variation in scales of disturbance.

3.3 Preliminary results: class-level fragmentation

At the class level, fragmentation metrics are calculated and returned for each class individually within a sampling cell; patches that do not belong to the same class are treated as NA values. In other words, class-level spatial statistics consider the relationships within individual patch clusters instead of across all patches.

We used the following categories which represent the number of years since a patch (or pixel) assigned to that class was last cleared: 0, 1-4, 5-11, 12-19, 20-49, 50+, pasture, and water. The groupings, which are slightly more aggregated than the “raw” survey results, were chosen during the classification workflow to minimize misclassification error rates. Pasture and water are excluded from this analysis to focus on the landscape patterns associated with forest succession.

While the automated classification generated by the trained random forest model is promising, there are artifacts of the process that must be manually corrected before we are confident in the results (Fig. 7). A supervised classification is being developed at the time of writing. Both rasters will be essential to guiding the final analysis.

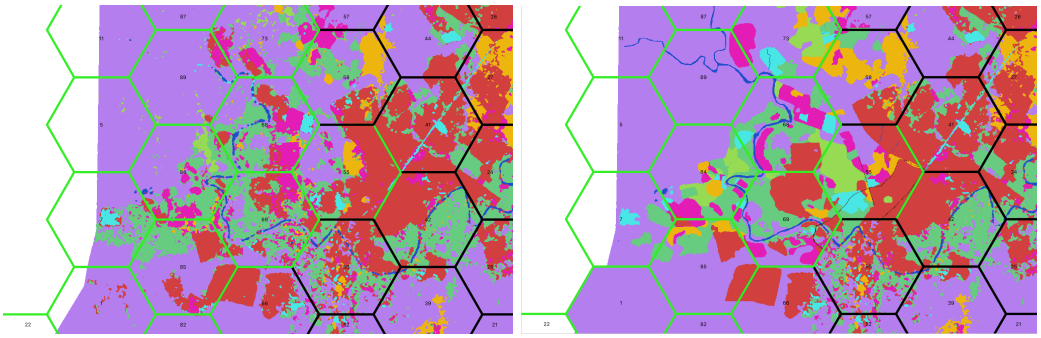


Figure 7: The automated classification (left) versus the supervised classification (right) where completed regions are outlined in green

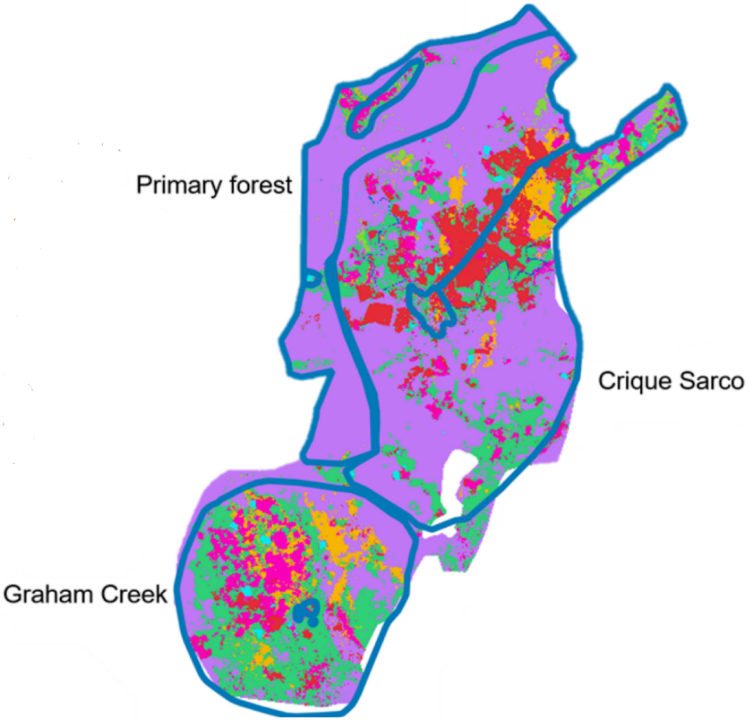


Figure 8: The full automated classification with regional polygons (Graham Creek, Crique Sarco, Primary forest) outlined in blue

For the purposes of this thesis, the fragmentation analysis is based on the automated classification, however, several metrics are highly sensitive to perimeter-area ratio and geometric complexity, and therefore we expect different results when the supervised classification becomes available.

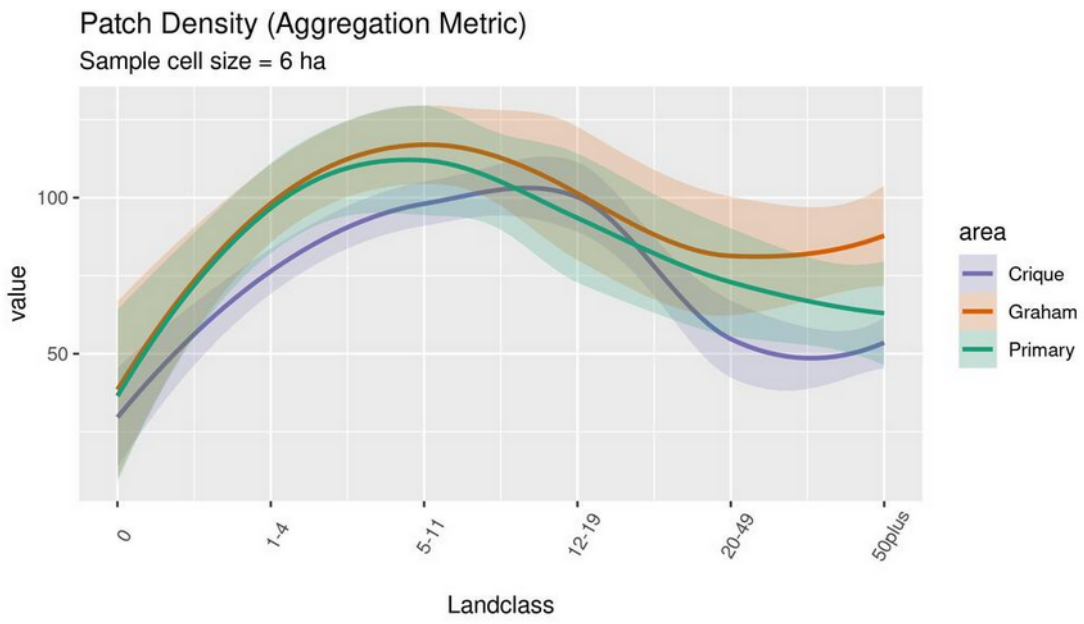


Figure 9: Mean patch density at Crique Sarco (purple), Graham Creek (orange), and the primary forest (green) with 95% confidence intervals (shaded) plotted across land-classes and filtered for majority non-NA (>70%) pixels

Fig. 9 plots the distribution of the patch density statistic in six forest regrowth classes (0, 1-4, 5-11, 12-19, 20-49, and 50plus years) and three sampling regions (Crique Sarco, Graham Creek, and Primary Forest [see Fig. 8]), defined as the number years of regrowth or recovery after clearing. Patch density is the number of spatially discrete "patches" per unit area, and in this figure it is averaged across the total area of each land-use class within each sampling region. The land-use classes are discrete ordinal categories, so the figure (and following figures) also includes a smoothed interpolation between classes to aid in interpreting temporal trends in the change in patch density as land use recovers from recent disturbance (0) to fully-regrown primary forest (50plus). Fig. 9 appears to show two areas of divergence: Crique Sarco and Graham Creek have higher patch density values on average in the 1-4, 5-11, and 12-19 year classes, with Graham Creek peaking slightly younger at 5-11 year classes and maintaining patch densities notably higher than Crique Sarco at 50+ year classes.

Crique Sarco is an older village and it was founded around 1910 (Downey 2015). This means the local population has been clearing its local forests for much of the 20th Century. In contrast, Graham Creek was settled around 1985 (Downey 2015). However, there are reports of an earlier

settlement in the 1960s that was subsequently abandoned (Downey 2015), which may help account for the higher patch density of older forests around Graham Creek. Both villages show signs of high recent clearing activity, with Graham Creek being slightly younger (maximum at 5-11 years compared to 12-19 years). In addition, in Graham Creek the patch density of the older age classes (especially >50 years) appear to diverge from the patch density of the primary forest. Because swidden cultivation tends to occur at an outward-expanding distance from the village center, as the patches close to the village regrow to a state where they can be cultivated successfully (usually 15 years and older), we would expect cultivation to return to these areas. This second stage of cultivation in the nearby fallow classes has likely occurred in Graham Creek. Across all land-use classes, Graham Creek tends to have higher patch density, which is a reflection of its younger age, the development of a significant fragmented mosaic of intermediate age classes, and fragmentation of the primary forest as cultivation expands outward. While few of these trends appear statistically differentiable (mostly overlapping confidence intervals), there are clear qualitative patterns.

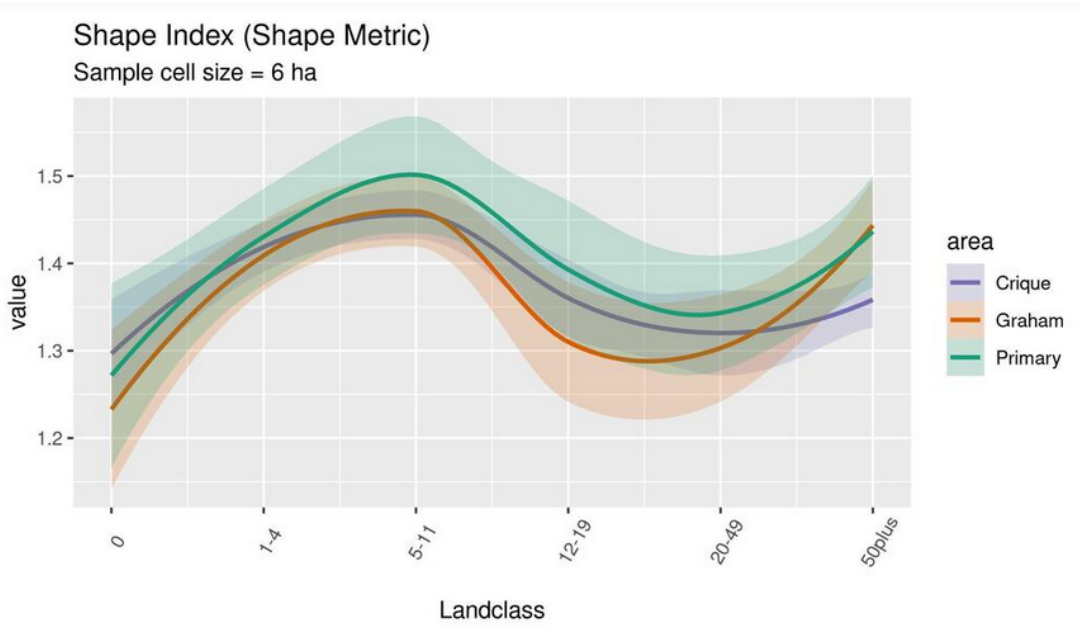


Figure 10: Mean shape index at Crique Sarco (purple), Graham Creek (orange), and the primary forest (green) with 95% confidence intervals (shaded) plotted across land-classes and filtered for majority non-NA (>70%) pixels

Next we consider the mean shape index at the class level. The shape index is calculated by dividing the patch perimeter by the patch perimeter of a square patch of the same size. The index equals 1 for square patches increases without limit as the patch becomes more geometrically complex, making the metric widely applicable in landscape ecological research (Forman and Godron 1986). Patton (1975) created the shape index to quantify habitat edge for wildlife species and as a means for comparing alternative habitat improvement efforts, such as wildlife clearings.

Because shape index measures the complexity of patch shape compared to a standard (square) shape of the same size, the metric does not depend on size like perimeter-area ratio. A preliminary interpretation of Fig. 10 would suggest that edge length is longer for all age classes in the primary forest, due to the more stochastic nature of natural disturbances such as tree falls, lightning strikes, and natural fires. . In the village farming areas, the shape index is lower across age classes, suggesting a higher degree of spatial organization due to human swidden clearing and cultivation activities. This could also potentially be an artifact of the smoothing algorithm used in the automated classification process.

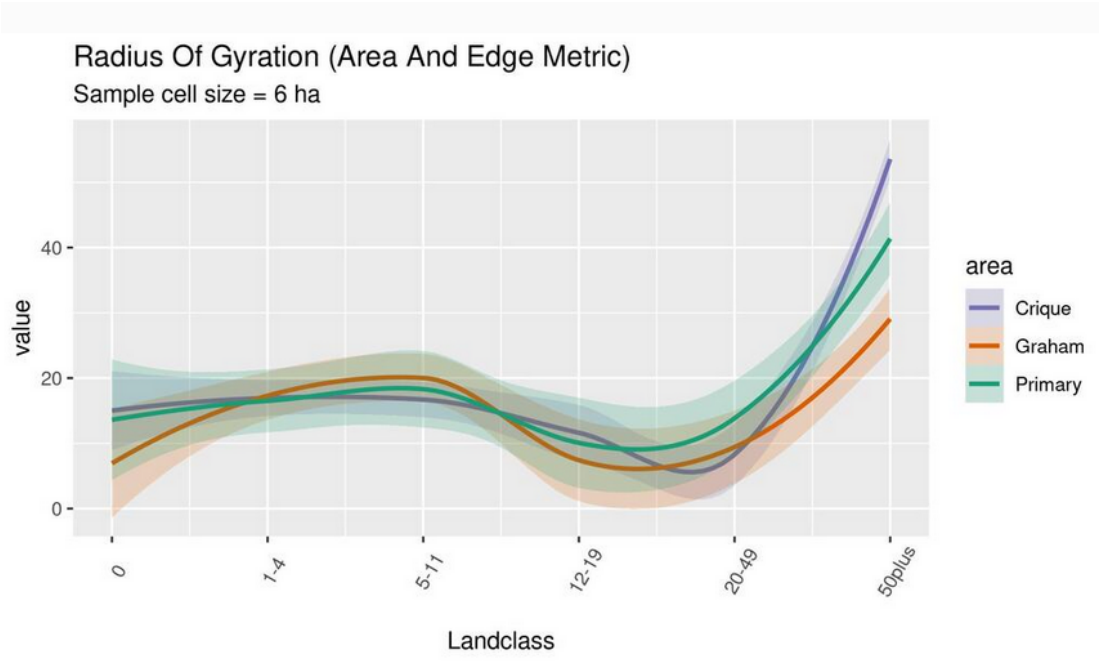


Figure 11: Mean radius of gyration at Crique Sarco (purple), Graham Creek (orange), and the primary forest (green) with 95% confidence intervals (shaded) plotted across land-classes and filtered for majority non-NA (>70%) pixels

Finally, consider the mean radius of gyration at the class-level, a measure of how far across the landscape a patch extends its reach, given by the mean distance between cells in a patch. In other words, the radius of gyration describes the average distance an individual can move within a patch before encountering another patch, given a random starting point. Larger or more elongated patches tend to have greater radii of gyration.

The distribution of the smoothed values is fairly indistinct between areas across class, but some divergence does occur with the primary forest (50+ years): a preliminary interpretation of Fig. 11 suggests that, on average, the size of the primary forest areas is larger in Crique Sarco than in the Primary Forest and Graham Creek, as larger average radii of gyration generally coincide with larger patch extent. However, this could be an issue with the polygon boundaries selected; Crique Sarco includes a large tract of primary in the southwest, making an analysis based on differences in the oldest class difficult.

One potential issue for future consideration is that these figures do not account for the amount of each class present in the polygon, e.g. there are very few, if any, 0 years and 1-4 years age classes in the Primary forest polygon. Cushman solved this by partialling out percent of the cover class in the landscape (Cushman 2008).

3.4 Preliminary results: landscape-level fragmentation

These statistics represent large-scale characterizations of the three different study areas. When generating fragmentation metrics at the landscape level, all patches are taken into account within each sampling cell regardless of class. Findings from these results can characterize the structure and composition of entire landscapes, rather than of specific landscape classes.

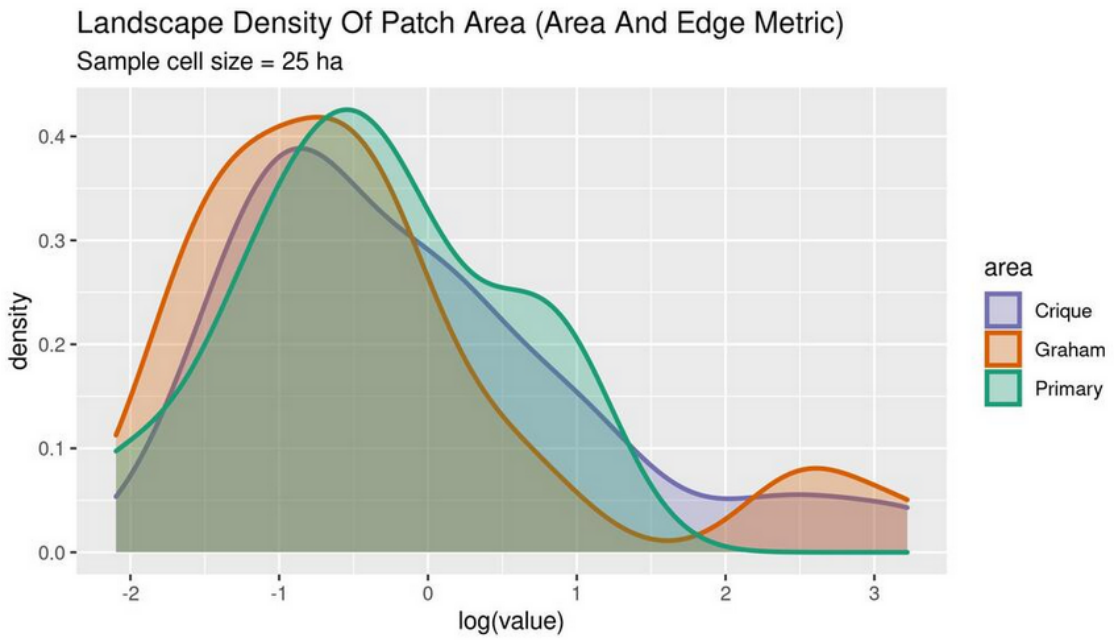


Figure 12: Density of mean patch area at Crique Sarco (purple), Graham Creek (orange), and the primary forest (green) plotted on the log scale and filtered by majority non-NA (>70%) pixels

At the landscape level, first consider the mean patch area. Fig. 12 compares the density distribution of the size of the patches using a log-normal plot. The primary forest has the largest mean patch size, and Graham Creek has the smallest. It is worth noting that the metric is highly sensitive to the size of the polygons that were used to analyze each area. In this case, Crique Sarco is a larger area, and contains some primary forest areas to the south-east of the village, which may affect its observed distribution.

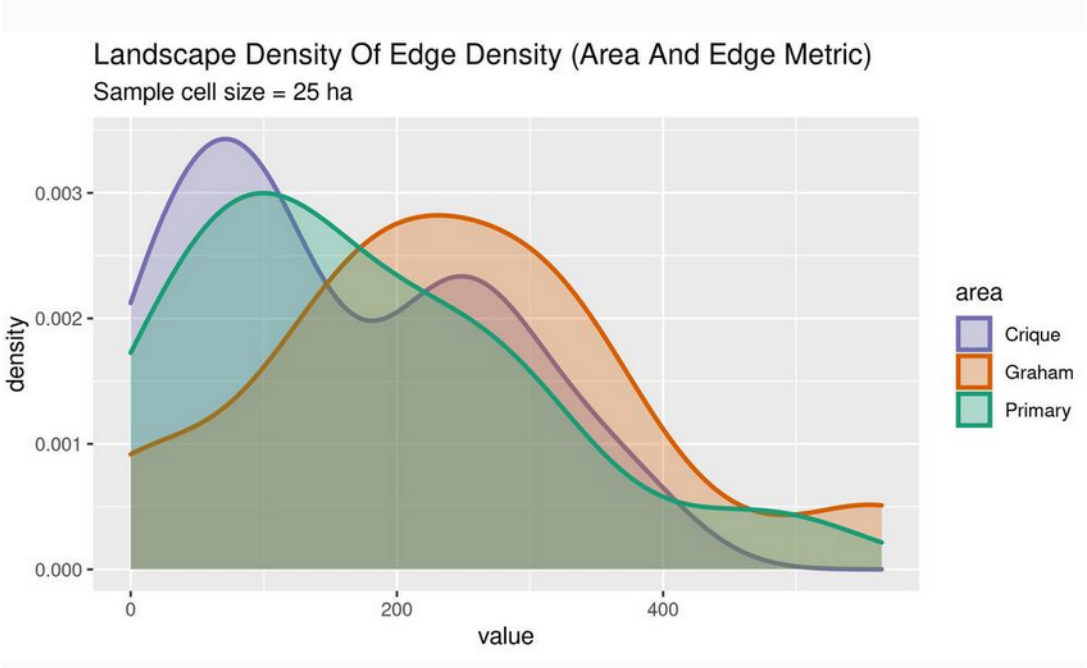


Figure 13: Density of edge density at Crique Sarco (purple), Graham Creek (orange), and the primary forest (green), filtered by majority non-NA (>70%) pixels

Next we will consider edge density, which is a direct quantification of the degree of fragmentation in each area. Edge density is a very important ecological variable because fragmented landscapes favors certain species, but may disadvantage others. This may be particularly important for analyses of prey species, which tend to favor habitat edges, or encourage deed dispersal of certain tree species.. Graham Creek can be seen to have the highest mean edge density. It is also interesting that the shape of these distributions vary: primary forest has a right skew, Graham Creek is omre normally distributed and Crique Sarco is bimodal.

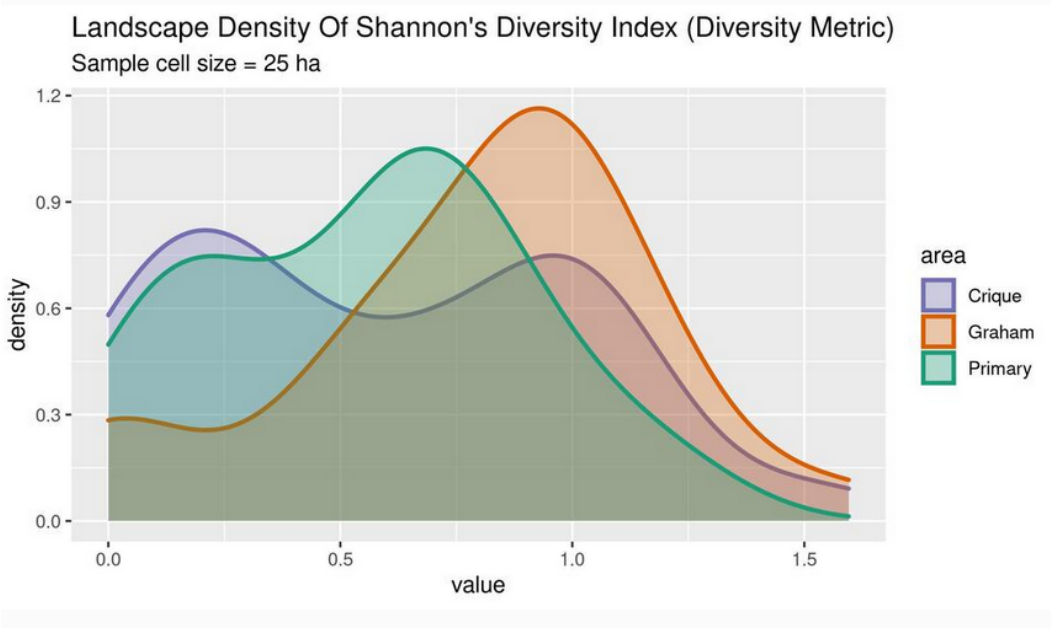


Figure 14: Density of shannon's diversity index at Crique Sarco (purple), Graham Creek (orange), and the primary forest (green), filtered by majority non-NA (>70%) pixels

Based on information theory, Shannon's diversity index can help describe the compositional makeup of the landscape. The value of the Shannon index for a particular sampling polygon represents the amount of "information" per individual (or patch type, in this case); larger values indicate a greater number of patch types and possibly greater evenness among patch types.

These concepts are based on the total number of patch types, which is derived from the number of land use classes (i.e., fallow classes), and their frequency distribution in each region. On their own, these statistics can provide information about patch diversity, but not the larger biodiversity of the system.

Graham Creek appears to have the highest Shannon diversity, followed by the primary forest and Crique Sarco, which exhibits a somewhat bimodal distribution. This is likely because Graham Creek is a very distinct, patchy mosaic, whereas the primary forest is structurally non-diverse. The density of the Shannon index in Crique Sarco may be somewhere in between.

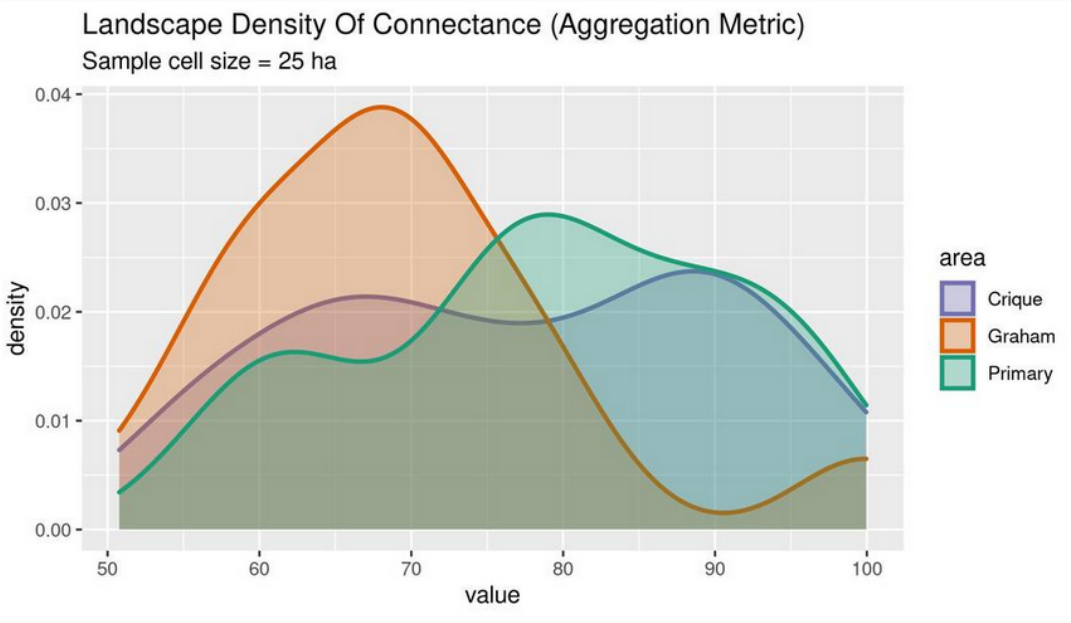


Figure 15: Density of Contagion Index at Crique Sarco (purple), Graham Creek (orange), and the primary forest (green), filtered by majority non-NA (>70%) pixels

Finally, consider the contagion index at the landscape level. Fig. 15 shows the contagion of the forests around Crique Sarco and Graham Creek, and the baseline values for the Primary forest. The Primary forest is a more contiguous block with greater contagion than the two villages, with Sarco appearing more like the Primary forest but still exhibiting some bimodal extrema. As with previous metrics, this can likely be attributed to the presence of primary forest tracks in Crique Sarco.

Overall, the fragmentation statistics provide a starting point for characterizing differences between the villages and the primary forest, which serves as a“control” point of reference. To fully describe the landscape and the impact of anthropogenic disturbances on local species, additional data and contexts are needed. This is why the biodiversity analysis, which incorporates hyperspectral and ground-truthed data, is necessary.

3.5 Parameter Tuning and Sensitivity Analyses

To get the appropriate 'cell' size to grid the classification map (and later the spectral species map) for distributed computing, many iterations over the tuning parameter were plotted for each metric until we could identify a preferred inflection point. This was optimized with respect to the scale of analysis -- i.e., small enough to make relevant comparisons across distinct regions (primary forest, both villages) but large enough disturbance patterns represent the collective practice of swidden by multiple farmers over time -- as well as previous research completed at this resolution and using these methods. A compromise must be made between the granularity of the data and the scale of interpretation, as we are interested in individual "species" identification but also larger collections of these species and the unique spatial structures that compose them.

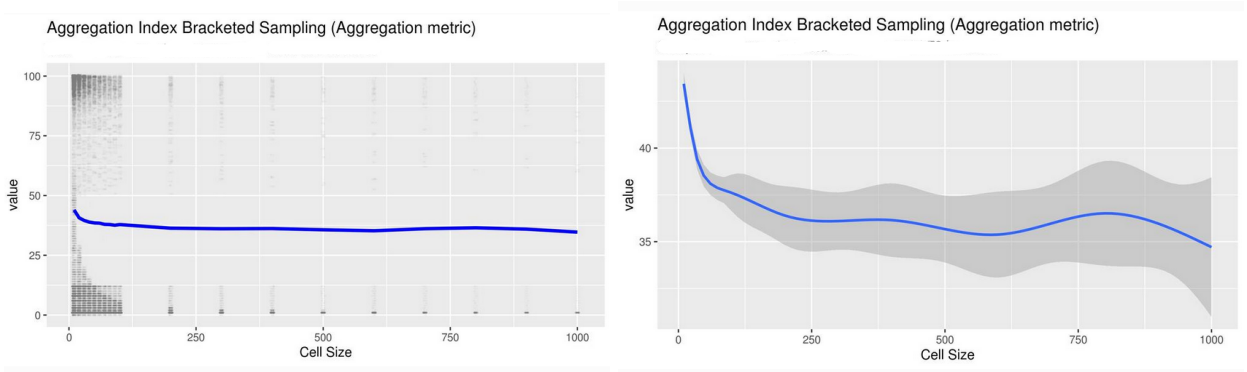


Figure 16: Plots of Aggregation Index distribution across various window sampling cell sizes (in meters) with smoothed mean loess function (blue) and 95% confidence intervals (shaded)

Note that the opacity of the points is extremely low for the sake of visualization, giving the appearance of non-normality in the density of the values. However, many of these values end up being excluded as outliers during thresholding (excluding sampling polygons which contain excessive [>30%] NA values), and the resulting density of points tends to appear less bimodal. Further caveats exist, which is why each metric will be interpreted on an individual basis.

Another important consideration are sampling polygons adjacent to the edge of the raster, which may contain a majority of empty, or NA, pixel values.

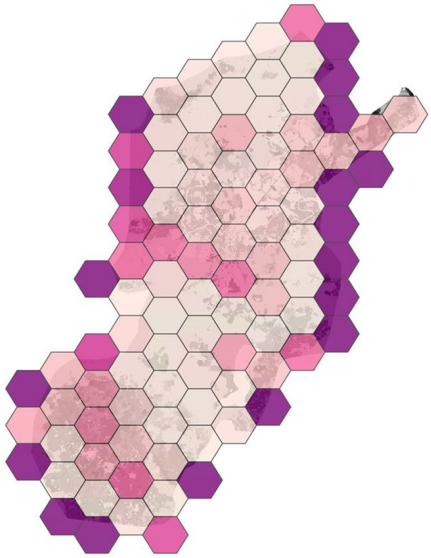


Figure 17: An example of edge effects skewing the interpretation of results if not excluded, where a darker purple hue represents a higher value of largest patch index (LPI)

Sampling window size had to be reconciled with the issue of edge effects: choosing a cell side-length too large (>1000m) would lead to more data exclusion on the basis of edge thresholding, limiting the amount of information captured in the statistics and force a broader scale of interpretation; comparatively, a very small cell side-length (<100m) could lead to less data exclusion on the basis of edge thresholding but severely limit the scale of interpretation. To remedy this, the relative difference in densities can be visualized before and after thresholding the data by how complete each cell is. This was tested over many cell sizes. Eventually, a heuristic of >70% was deemed significant enough to exclude extreme outliers but not so extreme that the edges of the map were unaccounted for.

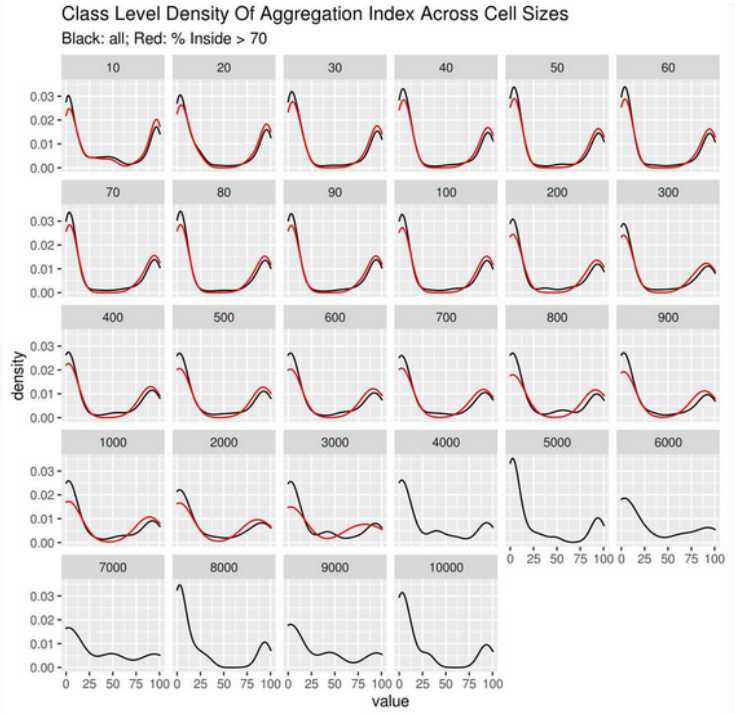


Figure 18: Class-level Density of Aggregation Index (value) across cell sizes (in meters) and data thresholds, where black indicates inclusion of all sampling cells and red is after the exclusion of cells containing <70% non-NA values.

After considering the impact of a 70% thresholding criterion across all statistics and cell sizes, a cell side-length of 250 meters was determined to be the ideal parameter value for calculations done at the class level, with the exception of Mean Euclidean Nearest Neighbor and Splitting Index (not shown), which exhibited convex behavior. At the landscape level, 500 meters was determined to be the ideal cell side-length, which equates to a sampling area of about 0.86 km^2 for a hexagonal grid.

3.6 Biodiversity analysis

To estimate biodiversity from our remote sensing data we use the biodivMapR package (Féret 2019), which also allows integration of ground-sampled species abundance data. This analysis uses the spectral variation hypothesis (SVH) and implements the R/biodivMapR workflow as defined in Féret 2019 (Fig. 19). Using the raw spectral data, first we perform a principal components analysis. Then, similar "species" are identified and clustered with k-means. We can

use field inventories (see Appendix 1) to validate the PCA and spectral species steps. Finally, the resulting raster is used to map α-diversity and β-diversity across the raster.

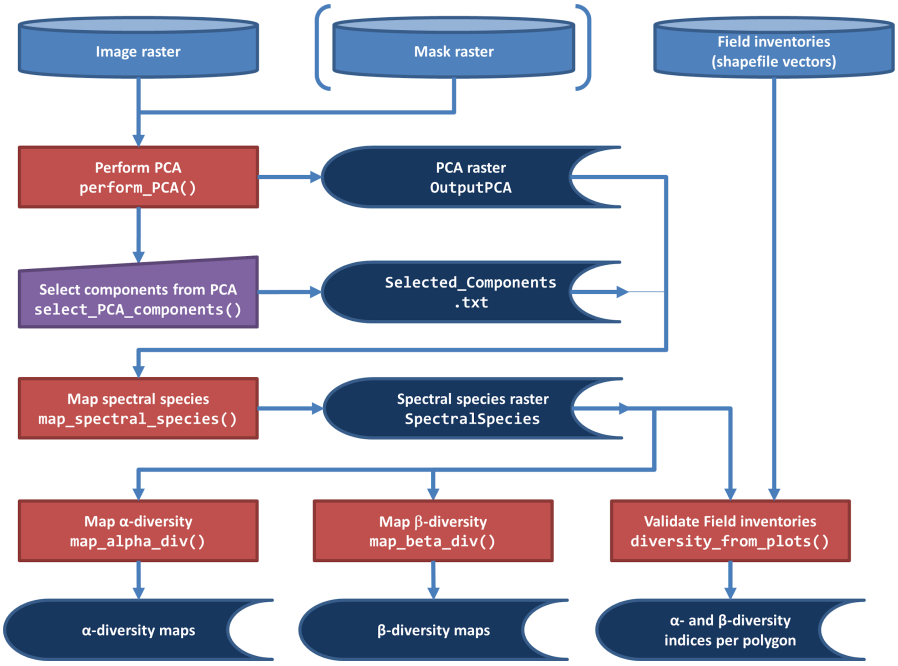


Figure 19: A flow chart of the computation of diversity maps using R/biodivMapR (Féret 2019)

The SVH states that spectral variance in remote sensing images correlates with habitat heterogeneity and species diversity (Oldeland et al. 2010). The results of various ecological studies show strong support for the SVH across a range of use cases. For example, richness and abundance-based diversity measures of vascular plants in central Namibia were found to have a strong relationship with spectral variation at both the $100m^2$ and $1000m^2$ scale (Oldeland et al. 2010). Hernandez (2012) used remote sensing data to assess the relative roles of productivity, habitat structure, and spatial variables in explaining observed patterns of α-and β-diversity. Rocchini (2007) examined the effects of scale on the to use remote sensing data to predict species richness and found that 4-band $3m^2$ QuickBird imagery exhibited a strong correlation (r=0.69), but that it varies with scale from about (r=0.20) at $100m^2$ to (r=0.50) at $10,000m^2$ (1ha). Odeland et al. (2010) examined the relationship between richness and abundance-based diversity measures from manual counts with spectral variability and found improvements in the Shannon Index at $100m^2$ and $1,000m^2$ scales of analysis (Oldelan et al. 2010). Overall, performance relies

heavily on the resolution or ground sampling distance of the remote sensing imagery and that performance improves at higher spatial resolutions.

The hypothesis this analysis will test is the Intermediate Disturbance Hypothesis (IDH), showing that disturbance at intermediate scales can potentially create an emergent landscape with high levels of structural diversity/habitat heterogeneity, thus enhancing the ecosystem. Applied to an anthropogenic landscape, the IDH posits that biodiversity is maximized when patterns of human alteration are significant but not overwhelming, which would lead to collapse; an intermediate level of anthropogenic disturbance creates the highest landscape productivity (Roxburgh et. al 2004, Bird 2015).

Using the high-resolution, multi-spectral spatial data and unsupervised learning, we can test for unique second-order structural dynamics. But first, several data management tasks were required to convert and re-scale the raw stitched multi-spectral images into the data needed to fit the specifications of version 1.7.0 of R/biodivMapR. To accomplish this, we ultimately corresponded directly with the author of the package, and he wrote additional functions to convert our spectral data into BIL format with the appropriate ENVI header information and rescale the data values to be only positive and standardized by the first and last bands, which exhibited the biggest outliers. The unusually large range of the data is an artifact of the data format used by the RedEdge sensor and the bespoke photogrammetric stitching process, which sought to minimize the loss of information.

The R/biodivMapR package offers robust radiometric filtering options to eliminate skew from non-vegetated pixels, shadowed or low-light pixels, and cloud-cover by choosing boundary thresholds of normalized difference vegetation index (NDVI) as well as near-infrared and blue spectra, respectively. However, since the stitching process already involves some level of radiometric filtering and the physical meaning of the resulting values has been transformed, it was difficult to define spectral thresholds based on published research and the software specifications. More work needs to be done to manually identify radiometric anomalies such as cloud cover, stitching artifacts, or smoke in the raster before a threshold boundary is accepted for any of these options. Therefore, a masking layer was not used for PCA generation in the workflow. This may change in the future, but should not significantly impact the results.

Prior to the principal components analysis, continuum-removal was applied to the raster to "reduce within‐crown variability caused by brightness variations, and effects of changing illumination conditions between flightlines" as described in section 2.3 of Féret 2019.

Three principal components were identified by the PCA and all were selected to proceed in the analysis, as three is a minimal amount for a landscape as diverse as the tropical forests of Belize. Testing was done at various levels of aggregation, including factors of 2, 3, 5, 10, 20, 30, and 36, the last of which was chosen to create a resolution close to that of the example raster in R/biodivMapR (Féret 2019).

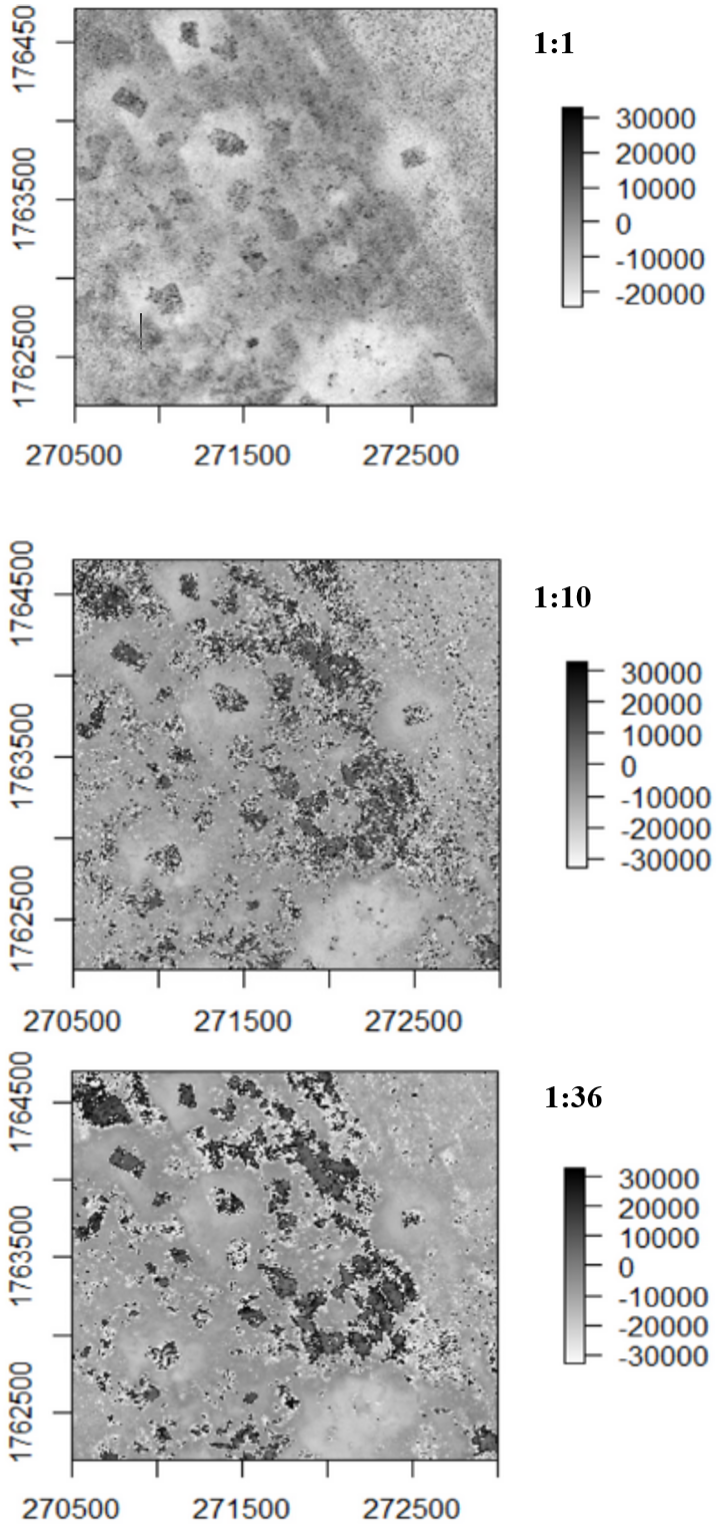


Figure 20: A subplot of the raster showing the effect of downsampling on the first spectral band of the original (top) at factors of 10 (middle) and 36 pixels (bottom)

However, no level of aggregation led to an increase in the number of principal components identified. Further manual comparison needs to be done to determine what resolution will optimize the selection of relevant features in the raster such as palms and large-crown trees, which can be easily identified (Fig. 21).

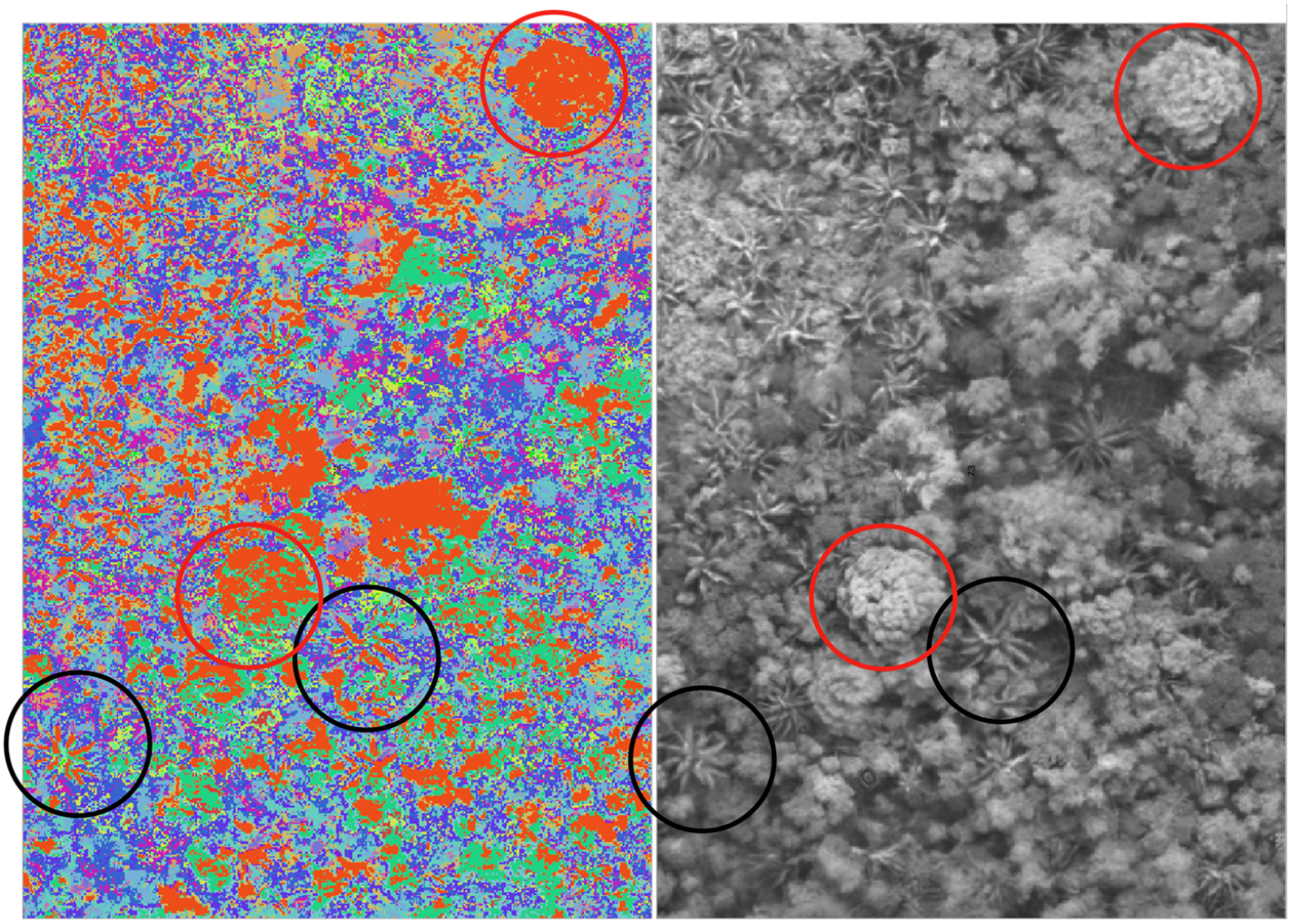


Figure 21: A visual comparison of the PCA (left), falsely colored in RGB, and the original raster (right) showing the fuzzy identification of palms (black) and large-crown trees (red)

Consider a subplot of the PCA and source raster above. The palms and large-crown trees, for example, are both some combination of two components at similar values. Furthermore, the distinct spatial pattern of the palms as identified by the PCA should be well-defined and consistent across individuals. Overall, the PCA is still failing to successfully identify species as individuals. As a result, species abundance counts could be inaccurate. Moreover, statistical

randomness and disorder may be artificially boosted beyond what we might expect. These shortcomings could possibly be attributed to the high resolution of the raster when compared to remote sensing data commonly available from satellites.

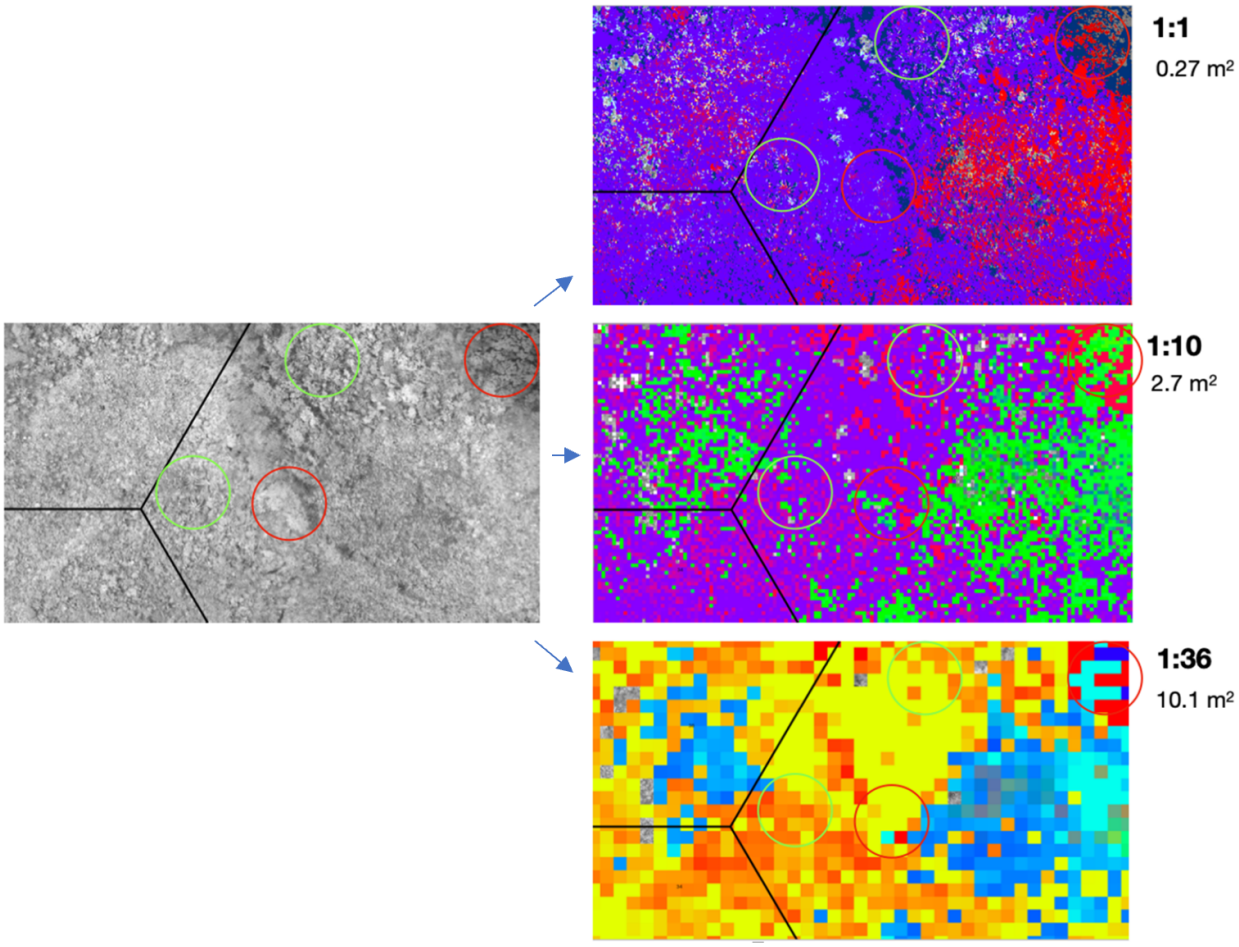


Figure 22: An example subplot (left) rescaled to illustrate the challenge of proper species identification (green and red circles) at various levels of aggregation (right), falsely colored by band

Ideally, each principal component will clearly identify varying aspects of diversity and structure within the larger landscape. Some preliminary results of the PCA show promising signs of differentiation among components, but again, further testing will be necessary for validation.

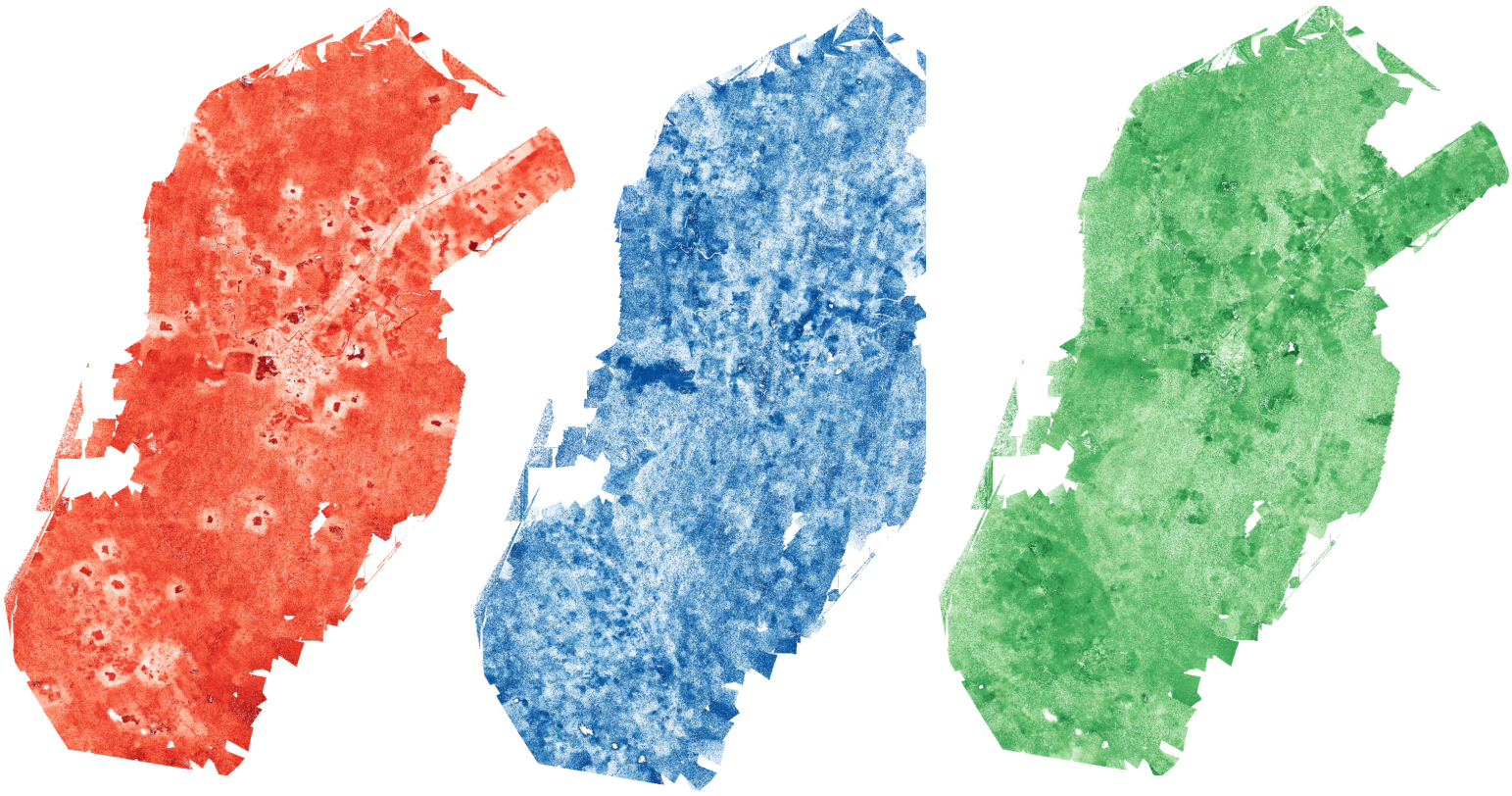


Figure 23: The three principal components, values falsely colored in quantiles of low to high intensity red, green, and blue, respectively.

A closer inspection allows for visual confirmation that the PCA is appropriately identifying differential components of the landscape. The patch below contains a range of landscape features, including a river as well as patchy clearings and neighboring fallow classes.

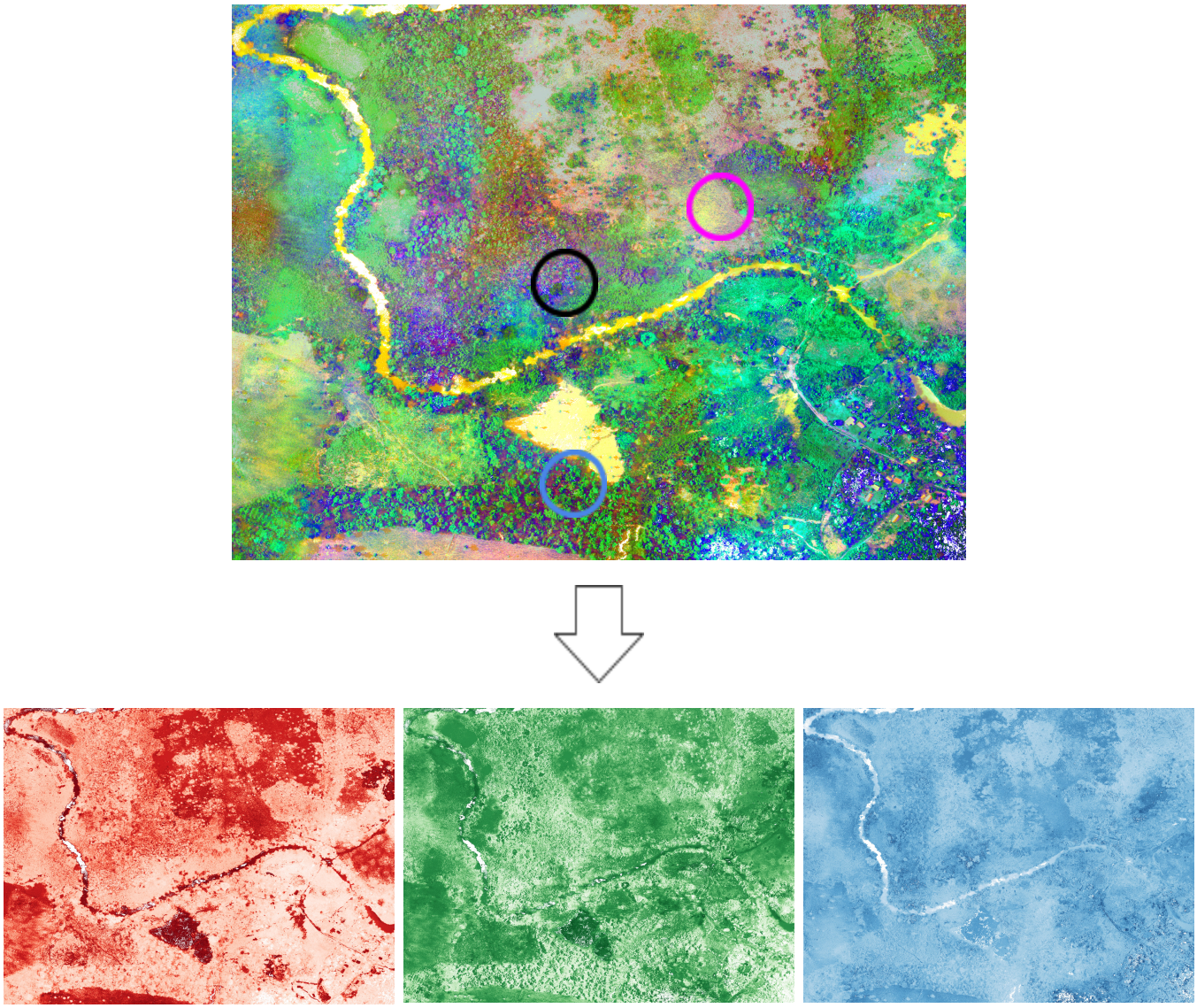


Figure 24: An example of the three principal components falsely colored with a quantile gradient from white to red, green, and blue, respectively (bottom) with the composite image above

Visually, each principal component appears to capture a different collection of features in the data (e.g. palms [black], mixed intermediate forest [blue], or recent clearings [purple]), although there may be some overlap. A more quantitative measure of the differential information captured within each eigenvector could improve confidence in the results of the PCA. Currently, a manual cross-reference with the source raster is in progress. It is likely that the default k-means parameters in the biodivMapR package (Féret 2019) will need to be modified to force additional principal components.

3.7 Spectral species identification

Due to the spatially complex nature of the drone scan area, we decided after the fragmentation analysis that square sampling cells did not achieve optimal accuracy in the estimated statistics and introduced more edge artifacts than necessary. The "spectral species" analysis was conducted using larger-scale, hexagonal cells. Prior to publication, we will include updated fragmentation metrics on the classified map so that any comparison across the two sampling frameworks is 1:1.

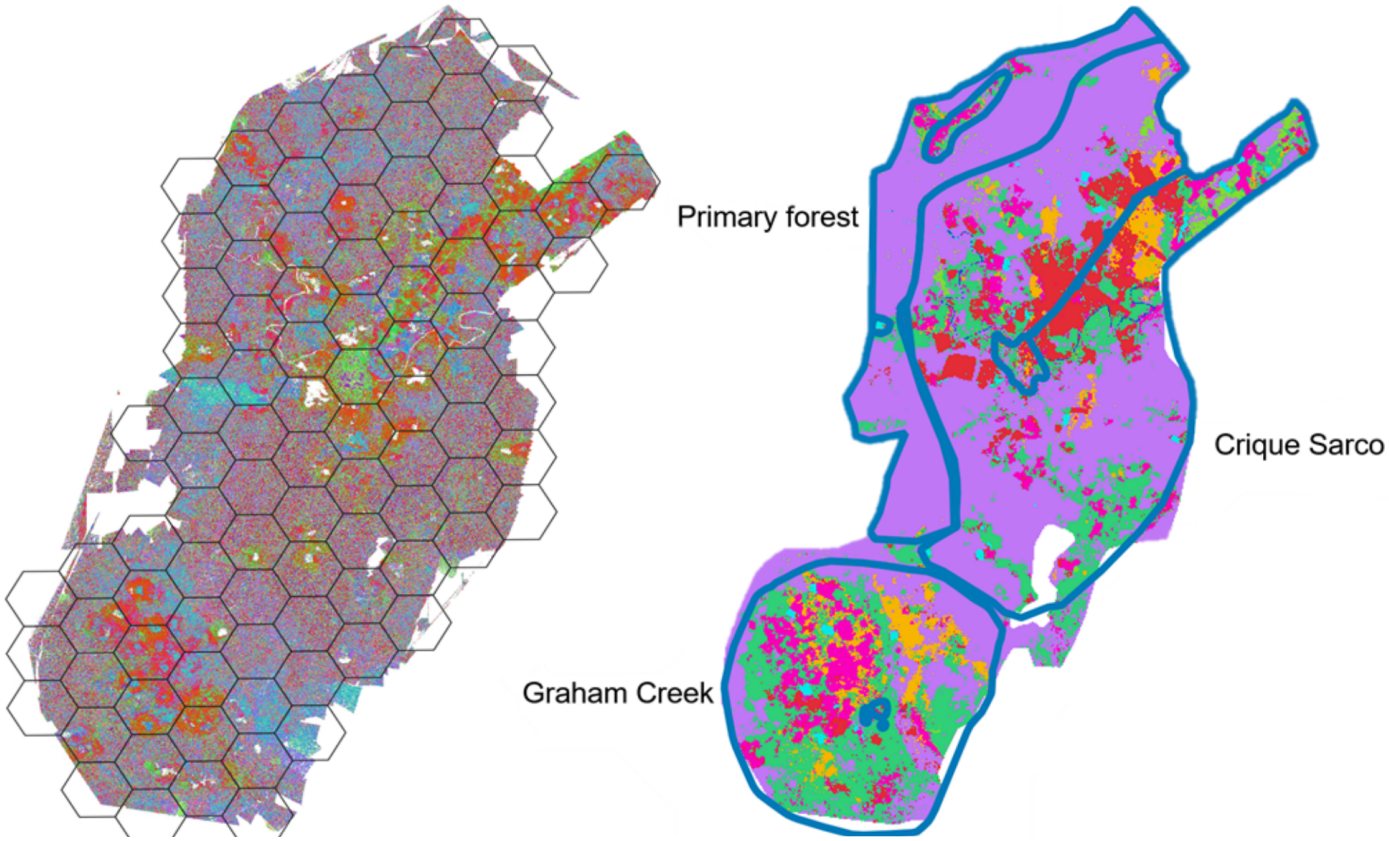


Figure 25: A map of the distribution of 20 spectral species identified by k-means clustering with sampling polygons overlaid (left) and the land classification map (right) with both villages and the primary forest region identified

Averaging over 11 partitions of the spectral PCA and 99 sampling polygons, each roughly 1km^2 in size and supplied as a polygon shapefile, 20 k-means clusters (or spectral "species") were

identified using R/biodivMapR (Féret 2019). A detailed explanation of the process can be found in section 2.5 of Féret 2019.

The number of partitions, *nb_partitions,* is automatically assigned by the k-means algorithm. However, a decision must be made about how many unique clusters, or *nbClusters*, k-means should identify. This determines the number of unique species possible across all sampling regions. A value too high will result in an overwhelmingly random assignment of "species" to each pixel; a value too low will be too patchy and aggregated, failing to capture the full variation in the landscape.

Féret and Asner (2014) noted that a tradeoff in performance can be made between biodiversity estimates and computational efficiency at 40 clusters. Féret (2019) sets the default *nbClusters* to 50 since the example raster is a diverse, tropical forest. However, Féret also notes that unpublished results from preliminary studies on "moderately diverse temperate forested sites showed optimal retrieval of Shannon index when using 20 clusters," suggesting that selection of the appropriate number of clusters may vary depending on the level of heterogeneity of the landscape (Féret 2019). Upon testing the full output of 20, 35, and 50 clusters, it became clear from qualitative inspection of the spatial clustering that 20 clusters, or "spectral species," was the optimal parameter setting for the purposes of our analysis. A simple visual comparison between the results of 1 of 11 partitions of the spectral species rasters generated using 20, 35, and 50 clusters also makes it clear that the raster becomes exceptionally random with more than 35 clusters, and patchier at 20 clusters (Fig 26). This could be due to the scale of our sampling regions with respect to the resolution of the raster, and further testing with downsampled rasters might be necessary to fully justify setting *nbClusters* to 20.

Figure 26: Spectral species at *nbCluster* = 20 (left), 35 (middle), and 50 (right), simply colored to distinguish larger area patchiness

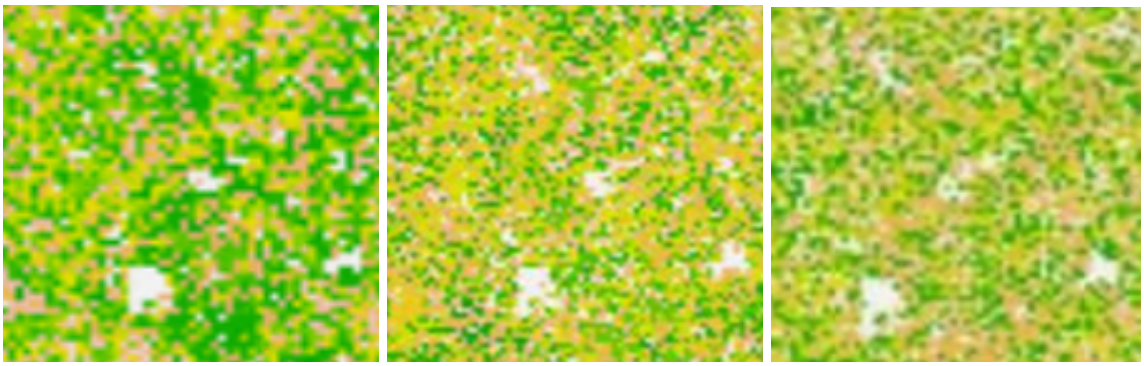


Figure 27: Spectral species at *nbCluster* = 20 (left), 35 (middle), and 50 (right), zoomed in

Note that there is no ordinal meaning to the color gradient, and data should be treated categorically; i.e., species "1" is interchangeable with species "20," and so on. Similarly, the choice of partition has no impact on the result of this visual comparison. The remaining 10 spectral species partitions for each *nbCluster* choice have been omitted because they exhibited the same patterns as those shown above (Fig. 26) and ultimately are averaged together.

3.8 Spectral species validation

To validate the spectral species raster and assist in understanding the composition of the principal components, we used the *compute_spectral_species_FieldPlots()* function in R/biodivMapR (Féret 2019). Polygons were manually curated by members of the HCLab from the GPS land survey data and verified using visual comparison to the spectral data in QGIS and field notes. These polygons were then treated as "ground-truthed" input for validating the principal components used to generate the spectral species raster; also, it is worth noting that this is the same dataset that was used in the automated landscape classification (Fig 3, pp. 11)

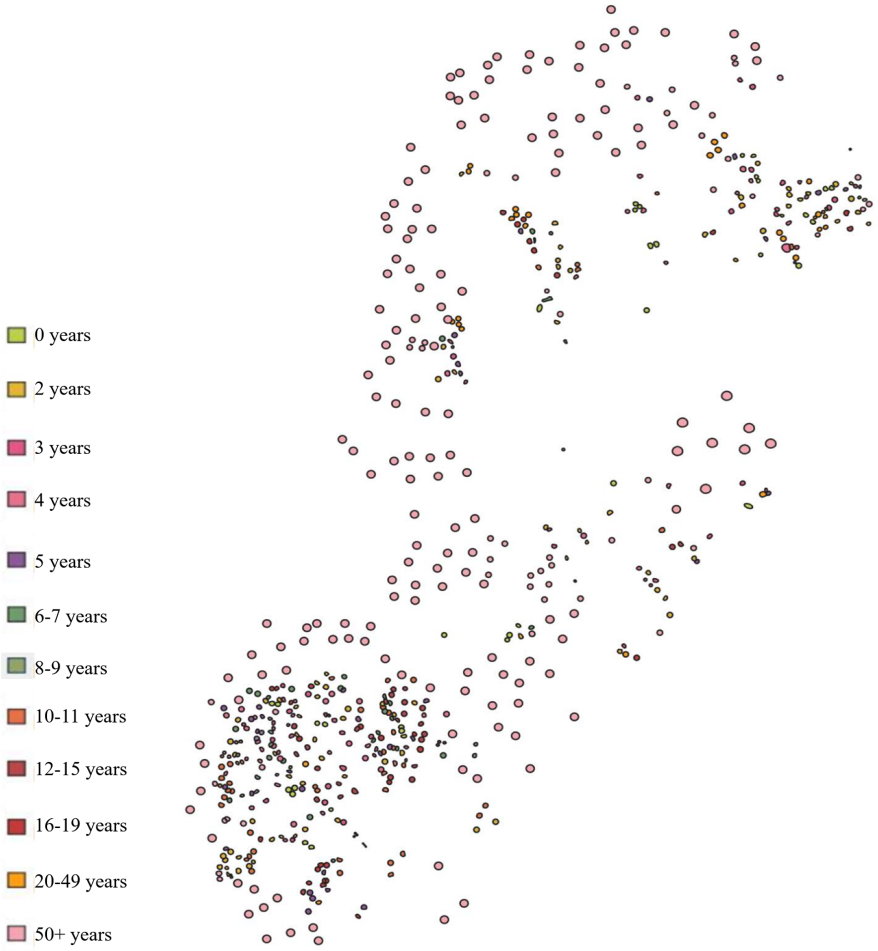


Figure 28: The study area with manually verified land class polygons from the 2018 survey. Land use classes shown in this map were grouped following the same criteria that was used in the automated classification.

We use 740 land-use training samples to validate the spectral species analysis. Fig. 29 shows land-use training samples clustering along the first two principal components, and overall agreement with four field plot samples where manual biodiversity estimates were counted by hand.

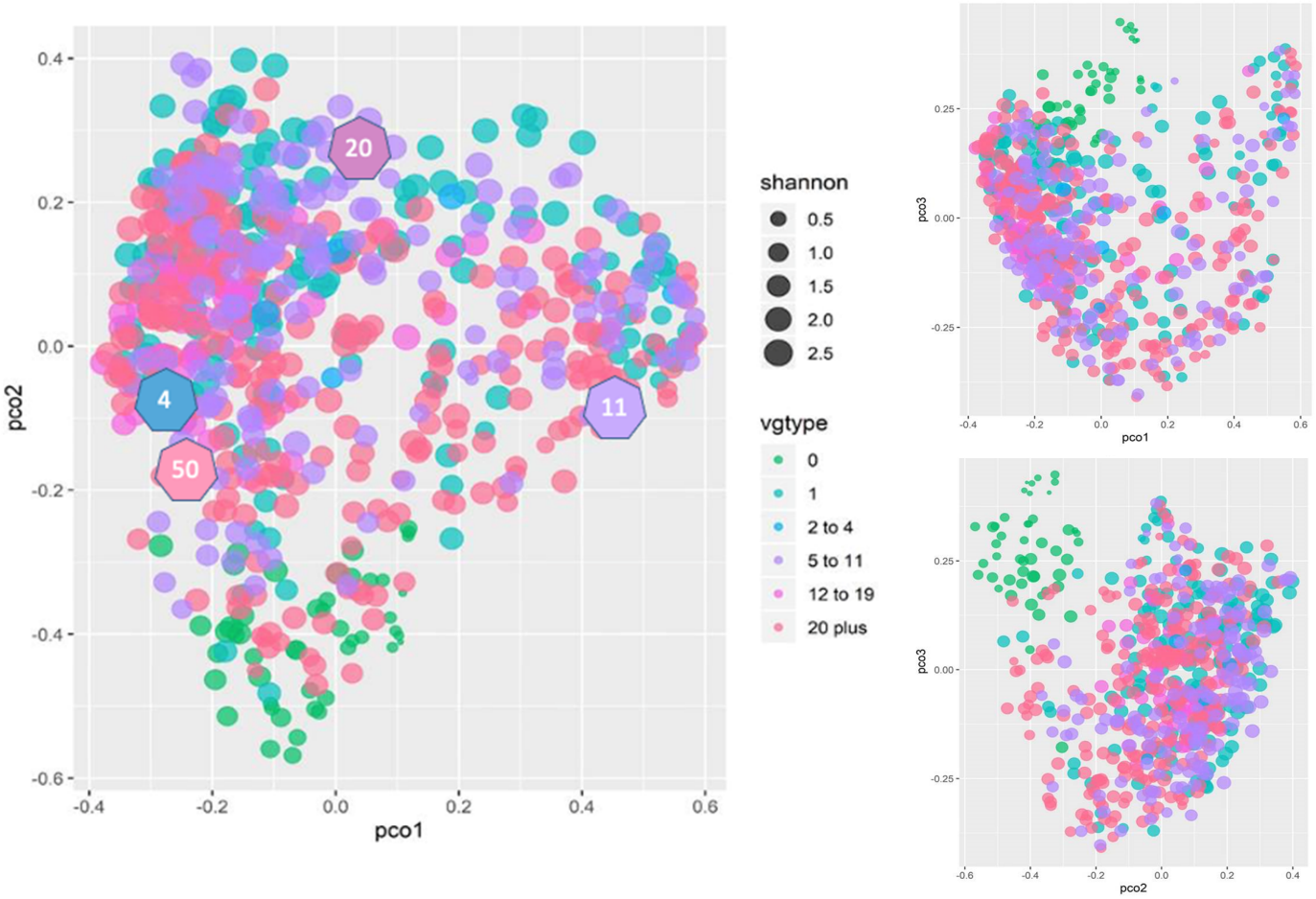


Figure 29: Validation plot of first two principal components from the spectral species analysis (left) as well as PC1 vs. PC3 (top right) and PC2 vs. PC3 (bottom right), with added polygons on the left plot representing where four ground-truthed biodiversity transects at four locations were located (see Appendix 1). The numbers in these polygons represent the approximate age-class of the forest patch where the transect was located

3.9 Spectral species α-diversity

Using the spectral species rasters as an input, we calculated diversity maps including metrics measuring both α-diversity and β-diversity. The computation of α-diversity metrics is based on

the distribution of clusters for a given window size, in pixels, over the whole image (Féret 2019). The Simpson index, a measure of species dominance, is also an option for the output of the α-diversity step, however we chose to use Shannon because of the focus on species diversity as a proxy for ecosystem enhancement. Depending on the spatial resolution of the image, type of ecosystem, and spatial scale of the expected diversity patterns expected, window size may also need to be adjusted. Selecting a window size that is too small would lead to lower range of variation and high noise sensitivity. According to Féret (2019), window sizes between 0.5 ha and 4 ha are usually compatible with field observations. For the sample raster, this would equate to 50–400 pixels of 10m resolution; for our data, this would correspond to about 1,786-14,286 pixels of 0.28m resolution unless aggregated.The diversity metrics are finally averaged across all (11) partitions of the spectral species raster, a methodology which is "analogous to signal averaging, increasing signal to noise ratio by replicating measurements" (Féret 2019).

Fig. 30 is a map of Shannon α-diversity across the spectral species raster.

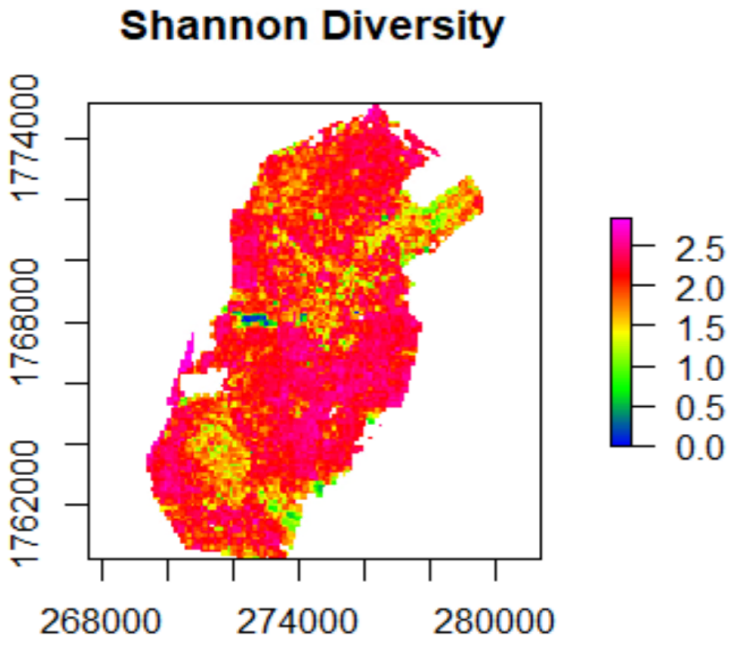


Figure 30: A α-diversity map of the study area colored by the average estimated Shannon index of that sampling window across all 11 partitions of the spectral species raster

3.10 Spectral species β-diversity

The process of calculating the β‑diversity maps is similar to that of α‑diversity, using a pairwise Bray–Curtis dissimilarity (BC) matrix obtained from the spectral species map. The *map_beta_div* function involves three main steps: first, on a limited subset of windows, ordination is used to define the projected 3D space from a dissimilarity matrix; second, dissimilarity between each window in the image and the subset of windows used in ordination is computed; finally, using inverse weighted distance and the coordinates of the three most similar windows from ordination, the coordinates of each window in the resulting 3D space are defined (Féret 2019). Both non‑metric multidimensional scaling (NMDS) and principal coordinate analysis (PCoA) are available as ordination methods, but NMDS tends to have a higher computational cost (Féret 2019), so we chose to use the default ordination method (PCoA) for our spectral species raster. Fig. 31 is a map of Shannon β-diversity.

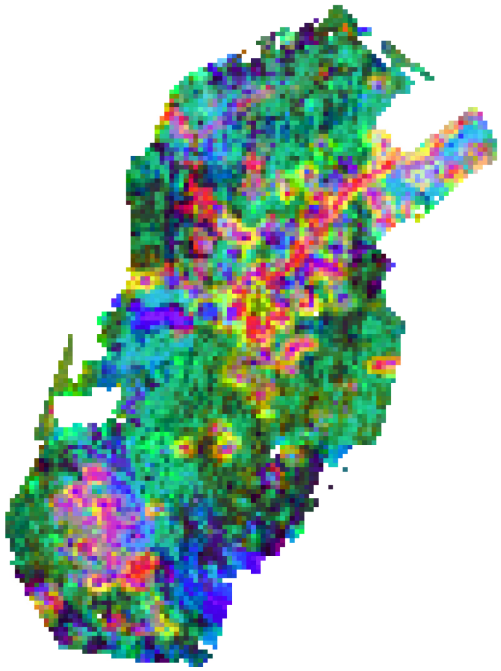


Figure 31: The bray-curtis dissimilarity of the identified principal components (PC1, PC2, PC3) of the spectral data overlaid in QGIS and colored red, green, and blue, respectively. Higher values are indicated by more vibrant color, while lower values are more transparent.

The results show a range of vegetation cover being identified by the individual principal components and will assist in characterizing the landscape's most important traits, but more manual cross-validation must be completed before we are ready to provide an ecological interpretation based on the β-diversity of the spectral species.

4. Connecting analysis

To test the IDH, we plot the relationship between landscape biodiversity and swidden disturbance, linking the results of the spectral species map to the results of the automated classification map. First, we identified the relevant statistics: within each sampling polygon, Shannon diversity was calculated from the spectral species raster and related to mean patch size, patch size coefficient of variation, edge density, and splitting index, which can identify landscape disturbance. The scatter-plots in this section include Bayesian credibility intervals and posterior predictions. The details of these models will be discussed in Section 5.

4.1 Preliminary results: class-level

We first considered using patch area as a measure of landscape disturbance in comparison to the (exponentiated) Shannon diversity of a given polygon. According to McGarigal (1995), the area of each patch comprising a landscape mosaic is one of the most important pieces of information (McGarigal 1995). Studies have found significant evidence that species richness as well as the occurrence and abundance of species are strongly correlated with patch size: for example, bird species abundance in Robbins et al. (1989).

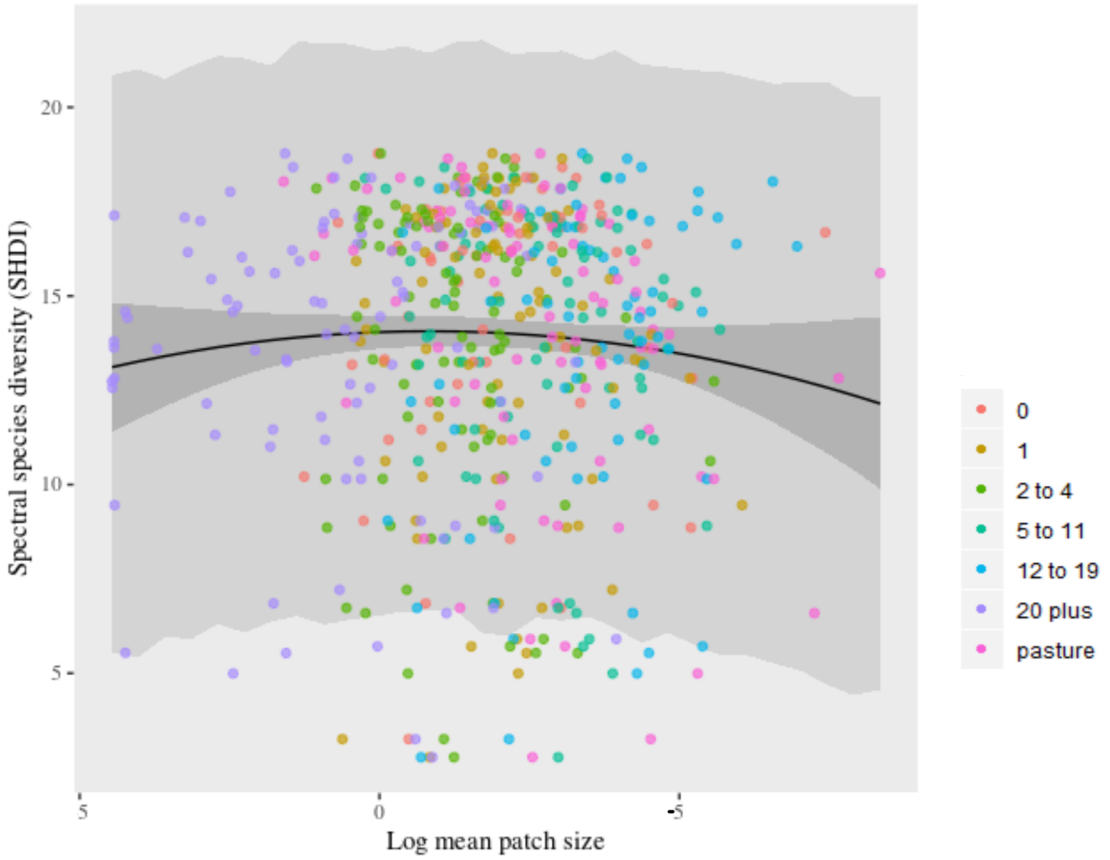


Figure 32: A scatter of exponentiated Shannon diversity of the spectral species vs. log mean patch size of the classified map, colored by age class. Bayesian posterior predictions (black), 95% credibility intervals of the fitted values (dark grey), and 95% credibility interval of the posterior predictions (light grey). Sampling cells containing >70% NA values were excluded. Complete modeling results and validation are presented in Section 5, and Appendix 2.2

The results in Fig. 32 can illuminate patch structure (e.g. many small patches or few large patches) at the class level. Note the reversed scale of the x axis. It is clear that the oldest patches, encompassed in the age class "20 plus," tend to be significantly larger, which we would expect given that the primary forest has experienced less disturbance than areas closer to the village. There is a slight downward, second-order curve to the points, which may be evidence that the IDH holds true in this landscape, but further testing is required to believe this signal. This is because at the class level, mean patch size is meaningful when comparing young classes to older forest classes. For example, we might calculate a ratio of the mean area of the youngest age classes (i.e. 0 years [cleared in 2018], 1 year, 2 to 4 years, 5 to 11 years, and 12 to 19 years) to

the area of the 20 plus years class, which would represent primary or secondary forest. Larger values of this ratio would indicate more disturbance.

4.2 Preliminary results: landscape-level

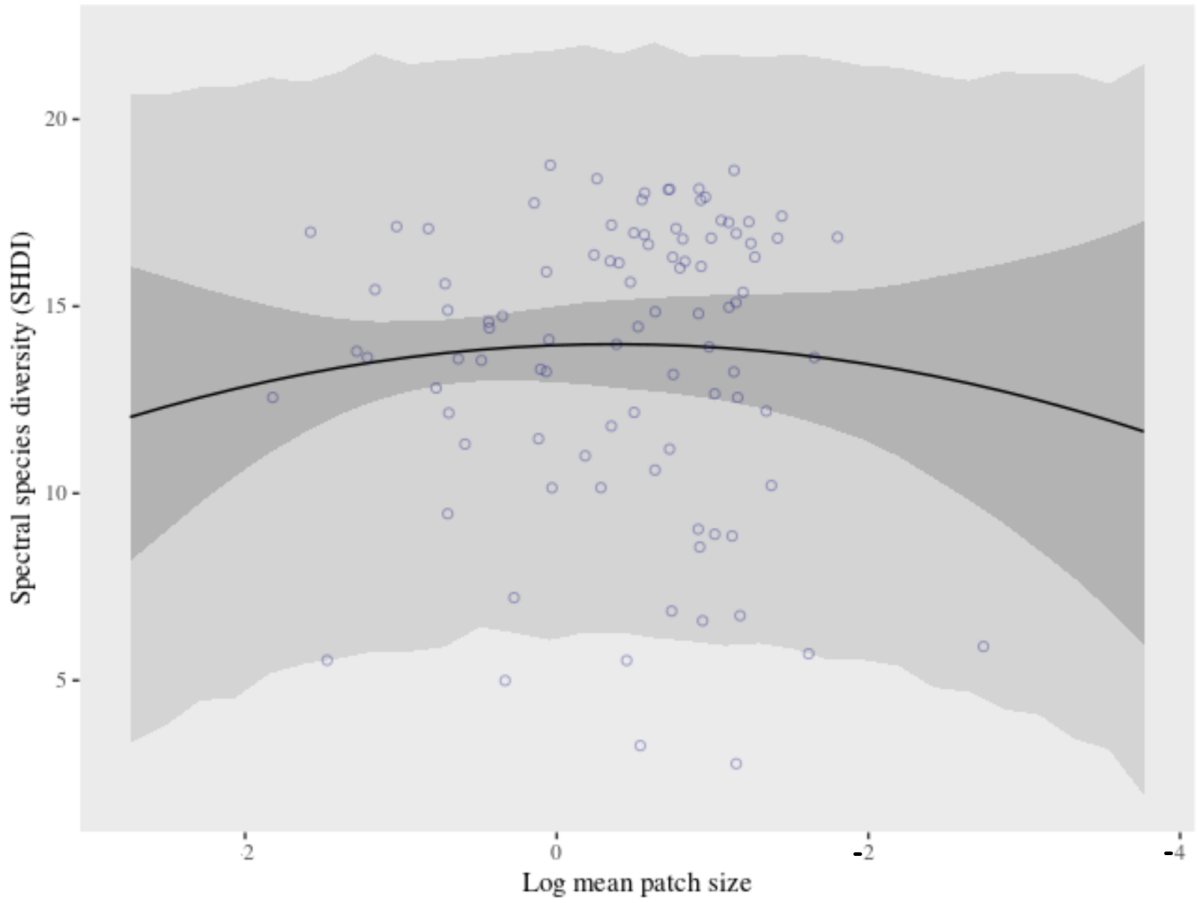


Figure 33: A scatter of exponentiated Shannon diversity of the spectral species vs. log mean patch size of the classified map at the landscape level. Bayesian posterior predictions (black), 95% credibility intervals of the fitted values (dark grey), and 95% credibility interval of the posterior predictions (light grey). Sampling cells containing >70% NA values were excluded. Complete modeling results and validation are presented in Section 5, and Appendix 2.3

Next, consider mean patch size at the landscape level. Again, increasing levels of disturbance (smaller mean patch sizes) are plotted to the right of the graph in Fig. 33. At the landscape level, the (log) mean patch size serves as an appropriate index for swidden fragmentation because a

primary forest polygon will tend to have more larger patches and few small patches regardless of fallow length, while a polygon near the village will contain more small patches.

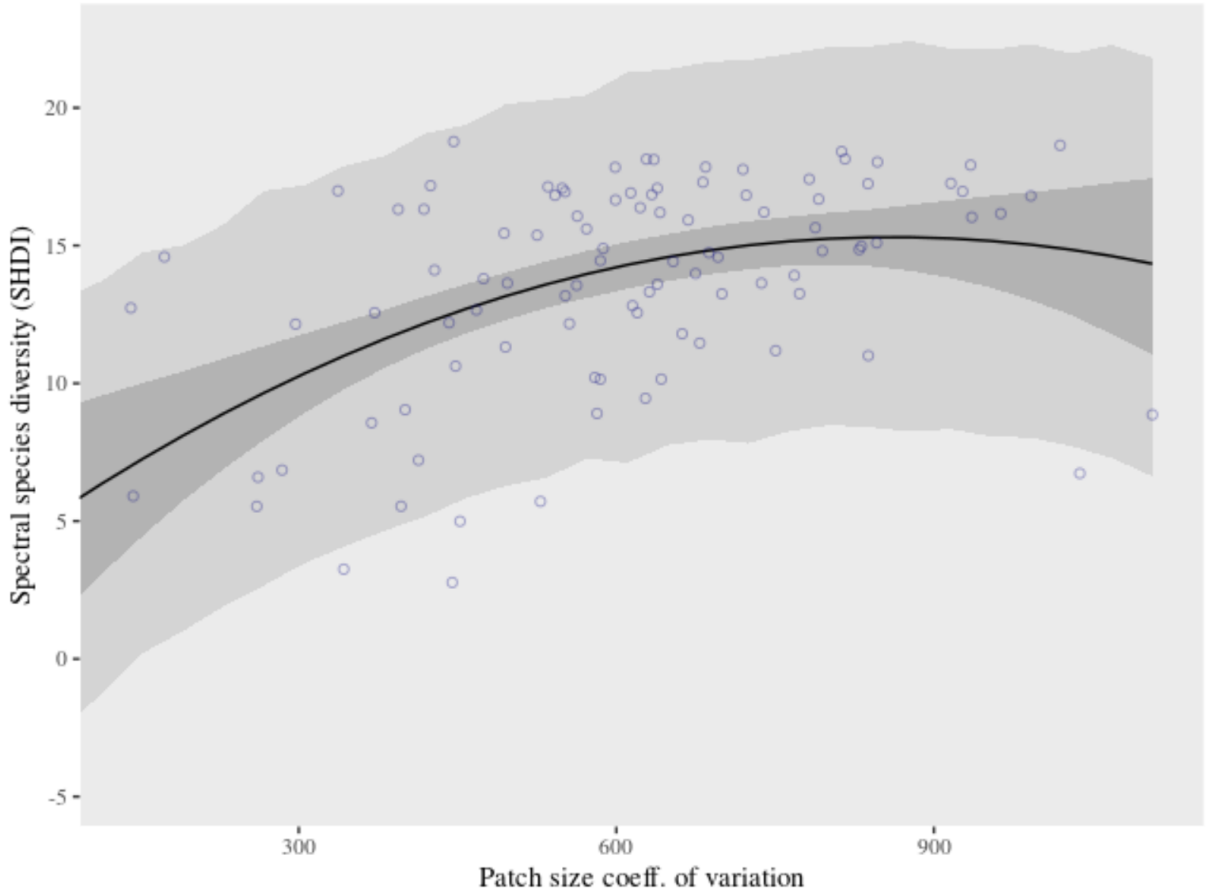


Figure 34: A scatter of exponentiated Shannon diversity of the spectral species vs. patch size coefficient of variation of the classified map at the landscape level. Bayesian posterior predictions (black), 95% credibility intervals of the fitted values (dark grey), and 95% credibility interval of the posterior predictions (light grey). Sampling cells containing >70% NA values were excluded. Complete modeling results and validation are presented in Section 5, and Appendix 2.4

Next, we considered the coefficient of variation of patch size, which can describe differences among patches in the landscape by measuring the relative variability about the mean (i.e., variability as a percentage of the mean), not absolute variability (McGarigal 1995). The coefficient of variation of patch size (AREA_CV) is scaled to the mean, making it easily

comparable across analyses. A greater AREA_CV indicates less uniformity in pattern and may reflect differences in underlying processes affecting the landscapes (McGarigal 1995).

We chose to calculate AREA_CV at the landscape level because the interpretation of AREA_CV at the class level is likely confounded due to the variation in older forest classes increasing with variation in younger classes, and both will be functions of patch size. Thus, AREA_CV will be more useful at the landscape scale. At the landscape scape, greater variability (higher AREA_CV) would be expected in more fragmented swidden landscapes (polygons) as the distribution of patch sizes increases across classes; less variability in patch sizes would be expected in natural landscapes like the primary forest. Because AREA_CV is a function of mean patch area, older and younger patches are directly comparable.

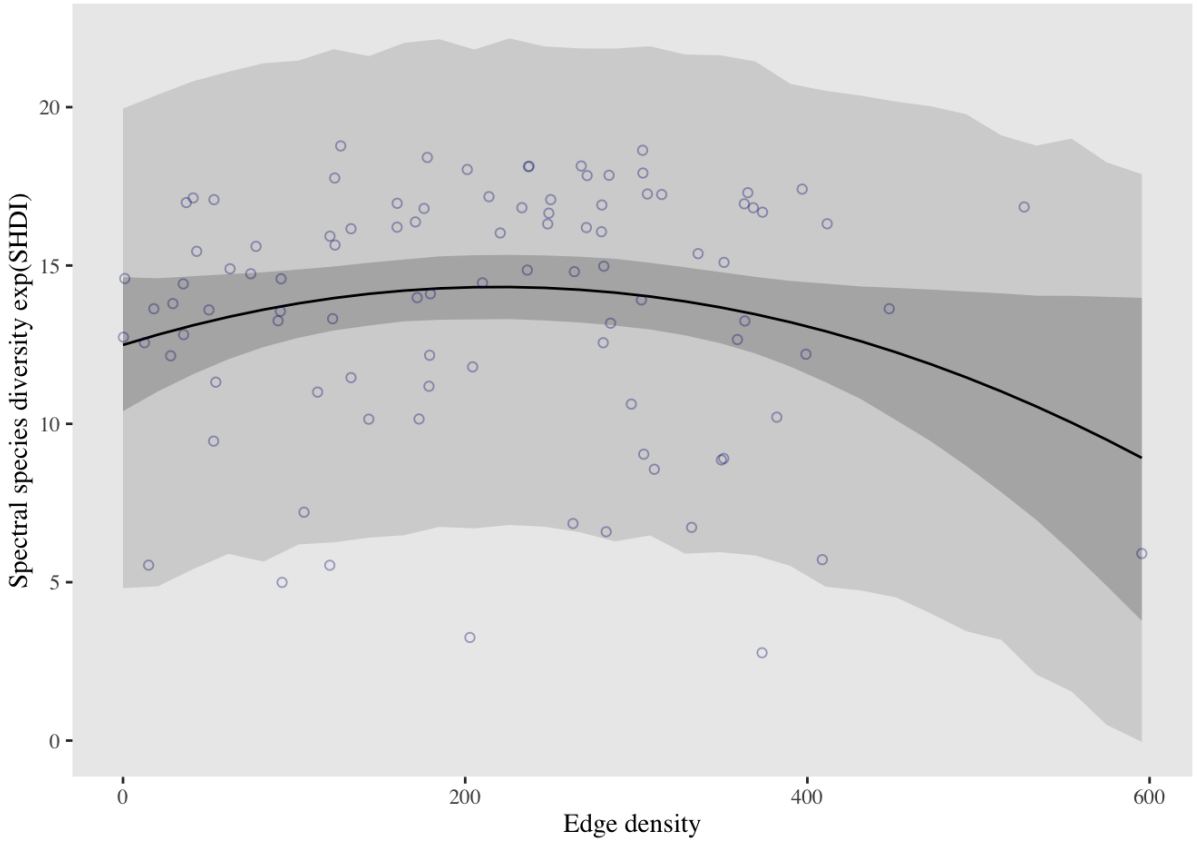


Figure 35: A scatter of exponentiated Shannon diversity of the spectral species vs. edge density of the classified map at the landscape level, Bayesian posterior predictions (black), 95% credibility intervals of the fitted values (dark grey), and 95% credibility interval of the posterior

predictions (light grey). Sampling cells containing >70% NA values were excluded. Complete modeling results and validation are presented in Section 5, and Appendix 2.5

Edge density, shown in Fig. 35 at the landscape level, describes the configuration of the landscape; an overall aggregation of classes will result in low edge density (McGarigal 1995). The edge density is also standardized to total landscape area. Similar to other metrics, there is a slight curvature that may suggest support of the IDH. We chose to use edge density over total edge (TE), another metric in this category, because the edge density will be calculated based on the proportion of the polygon that is filled, making it an ideal metric for polygons that lie on the edge of the raster.

At the class level, interpretation of edge density would have to be done as a comparison or ratio between young and old age classes, but since it is a function of perimeter (which correlates with mean patch size), this will likely not be as informative as comparing a ratio of mean patch size between classes (above). For this reason, edge density will be more useful at the landscape scale.

At the landscape scale, edge density assesses the level of fragmentation (number or edges) in each polygon. In undisturbed forest patches, edge density will tend to be lower, and in disturbed swidden patches it will be higher. However, as human disturbance increases even higher, edge density could increase as ecosystem dynamics transition to a cleared state, patches merge, and forest regrowth dynamics falter. The clearest example of this is cattle pasture class. It is also worth noting that edge density will tend to be higher when analyzing the unsupervised classifications, because of the spatial artifacts introduced by the smoothing algorithm, however this error will propagate evenly leading to systematic error across all classes.

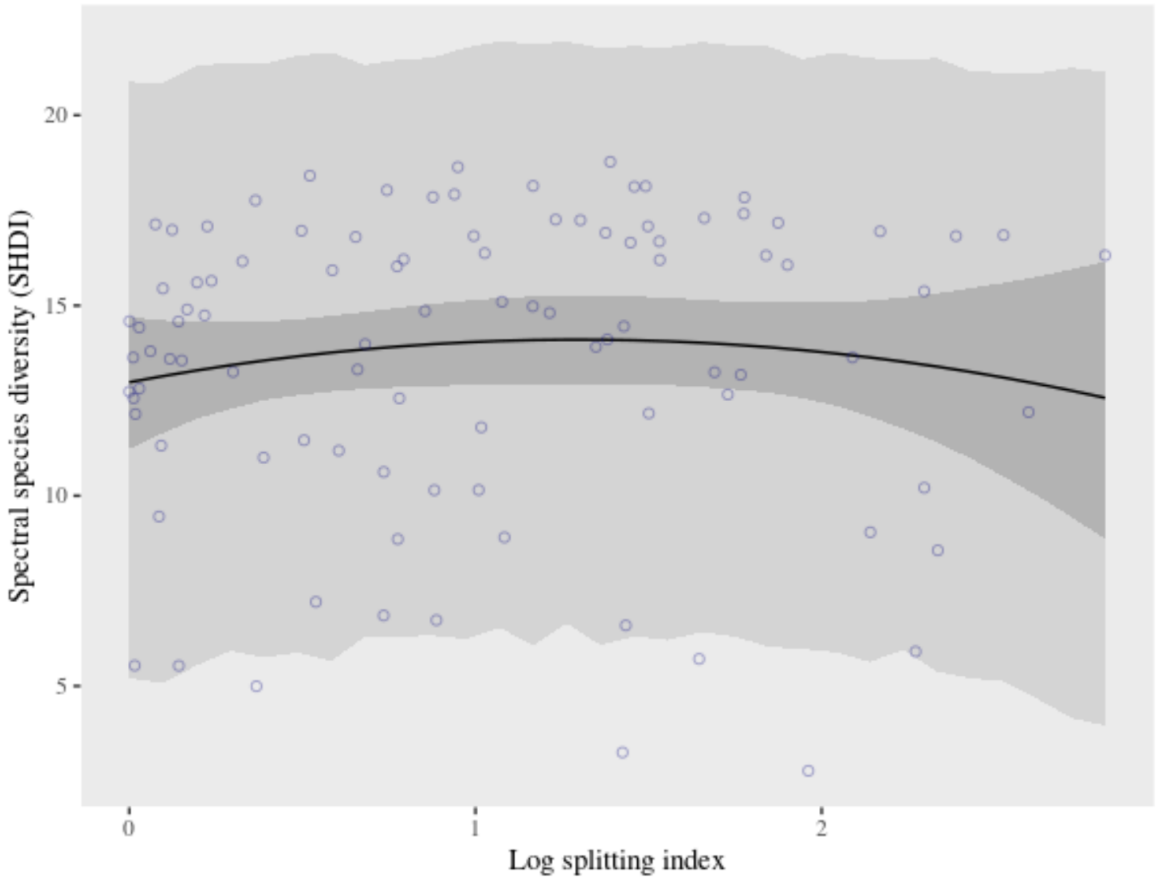


Figure 36: A scatter of exponentiated Shannon diversity of the spectral species vs. log splitting index of the classified map at the landscape level. Bayesian posterior predictions (black), 95% credibility intervals of the fitted values (dark grey), and 95% credibility interval of the posterior predictions (light grey). Sampling cells containing >70% NA values were excluded. Complete modeling results and validation are presented in Section 5, and Appendix 2.6

Lastly, splitting index was considered at the landscape level. The splitting index describes the number of patches if all patches in the landscape were divided equally such that this new configuration leads to the same degree of landscape division as obtained for the observed cumulative area distribution (McGarigal 1995).

At the landscape level, splitting index can be used as a general index of the patch fragmentation. One advantage of this metric lies in its insensitivity to the omission or addition of very small patches. This is particularly important in our analysis when using the automated classification,

which tends to have artificially small spatial clusters that result from the processing. The resulting statistic is a positive integer, so it will be intuitive to interpret as a disturbance metric.

5. Bayesian statistical modeling results

We conducted a Bayesian analysis that models the 2nd-order relationship between forest fragmentation and ecosystem biodiversity and which calculates the posterior probability of maximizing biodiversity at intermediate scales of disturbance. These models explore whether forest resiliency and health coincide with intermediate levels of human disturbance, and provide preliminary quantitative evidence showing that swidden cultivators may act as trophic mediators in Belizean tropical forests.

We fitted 2nd-order polynomial Gaussian models with relatively loose regularizing priors to the simulated data and five pairs of variables from the landscape fragmentation and spectral species analysis (see Equation 1). Models were run with 4 chains and for 6000 after a 2000 iteration burn using R/brms (Bürkner 2017). All models produced convergence plots similar to Fig. 37 (see Appendix 2.1-2.6). Convergence indicates that the MCMC estimates have reached an equilibrium and that the parameter estimates can be trusted. Table 3 contains coefficient summaries of the five fitted Bayesian models. Full statistical summaries and convergence plots are available for each model in Appendix 2.

$$shdi_i \sim Normal(\mu_i = \alpha + \beta_1 ed_i + \beta_2 ed_i^2, \sigma)$$

$$\alpha \sim Normal(50, 100)$$

$$\beta \sim Normal(0, 10)$$

$$\sigma \sim Cauchy(0, 1)$$

Equation 1: a Gaussian model relating edge density to Shannon diversity at the landscape level with regularizing priors for intercept, slope, and variance

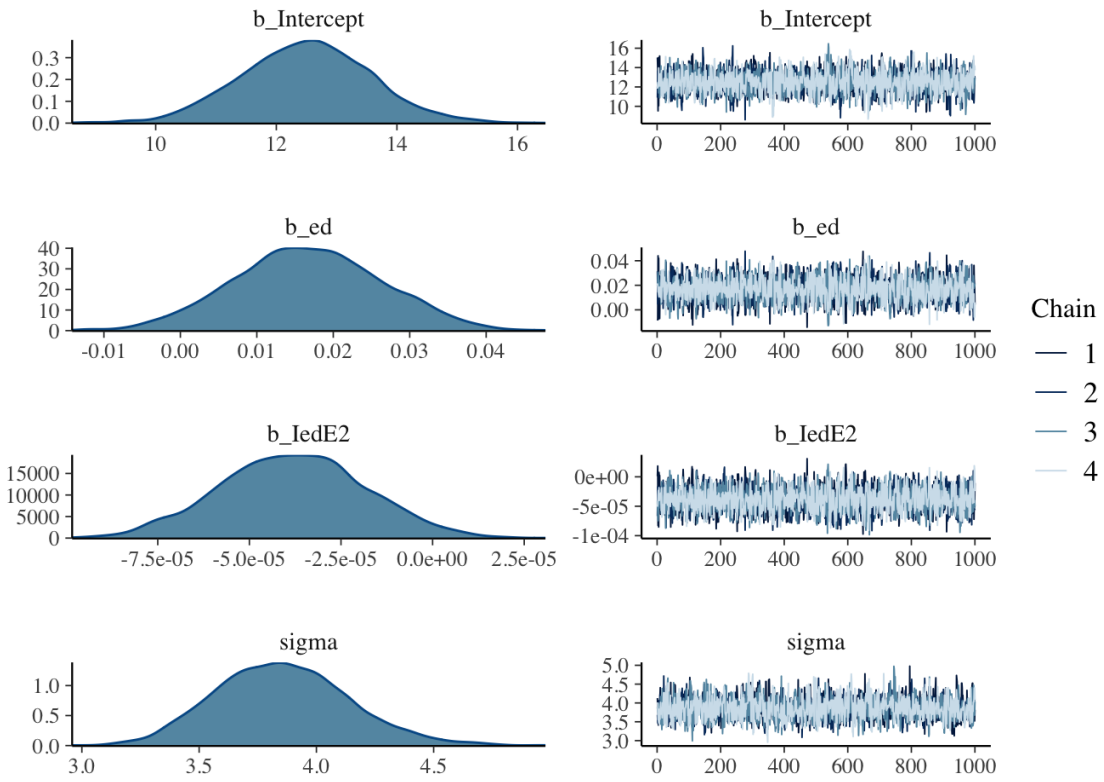


Figure 37: Convergence plots of the 2nd-order Bayesian model relating Shannon diversity to landscape edge density

***Model* [scale, Figure]**	**Intercept**		**Variable**		**Variable^2**	
	Est. [*error*]	**95% CI**	**Est. [*error*]**	**95% CI**	**Est. [*error*]**	**95% CI**
Simulation [*Fig. 1*]	-2.25 [*0.01*]	(-2.27,-2.22)	0.13 [*0.00*]	(0.13,0.13)	0.00 [*0.00*]	(0.00,0.00)
exp(SHDI)~1+log(area_mn)+ log(area_mn)^2 [*class-level, Fig. 32*]	14.04 [*0.22*]	(13.60,14.49)	-0.06 [*0.10*]	(-0.26,0.15)	-0.03 [*0.02*]	(-0.08,0.01)
exp(SHDI)~1+log(area_mn)+log(area_mn)^2 [*landscape-level, Fig. 33*]	13.96 [*0.52*]	(12.97,15.00)	0.15 [*0.42*]	(-0.69,0.98)	-0.20 [*0.22*]	(-0.64,0.23)
exp(SHDI)~1+area_cv+ area_cv^2 [*landscape-level, Fig. 34*]	3.42 [*2.45*]	(-1.37,8.12)	0.03 [*0.01*]	(0.01,0.04)	0.00 [*0.00*]	(0.00,0.00)

exp(SHDI)~1+ed+ed^2 [*landscape-level, Fig. 35*]	12.49 [*1.07*]	(10.40,14.63)	0.02 [*0.01*]	(0.00,0.04)	0.00 [*0.00*]	(0.00,0.00)
exp(SHDI)~1+split+ split^2 [*landscape-level, Fig. 36*]	12.98 [*0.86*]	(11.23,14.70)	1.72 [*1.72*]	(-1.73,5.13)	-0.66 [*0.71*]	(-2.06,0.71)

Table 3: Summary of fitted Gaussian 2nd-order polynomial models of spectral diversity for simulated data and for five pairs of landscape variables

Coefficient estimates for polynomial models are difficult to interpret (McElreath 2020:112). For example, in model exp(SHDI)~1+ed+ed^2 (Fig. 35) the intercept tells us the expected value of exp(SHDI) when edge density is at its mean value, but because it is a non-linear polynomial regression it will not necessarily match the mean value of exp(SHDI) from the sample. In Table 3, the variable and variable^2 coefficients are the linear and squared components of the parabola; however ecological interpretation of these values are not immediately apparent. Thus, McElreath's recommended approach is to plot the fitted values and the predictions from the posterior distribution, as shown in Figures 32-36.

For example, Fig. 32 and Fig. 33 show mean patch size in relation to canopy tree species diversity. At the class level (Fig. 32), this is particularly helpful in comparing patches of younger species to those of older species. It is clear that species diversity is maximized where a mix of age-classes fall while the oldest (20plus) and newest (pasture) points, which lie at the ends of the plot, are noticeably lower in diversity. Fig. 33 investigates mean patch size at the landscape level, showing highest canopy tree species diversity at intermediate ranges (~0.3 hectares to ~2.7 hectare) compared to the full range (~0.05 hectares to ~7.3 hectares). Fig. 34 shows the patch size coefficient of variation producing the highest canopy tree species diversity at intermediate to higher levels (800) compared to the full range (0-1200), where a value of 0 would indicate complete homogeneity among patch sizes, independent of age. We would expect less variability in patch size within natural landscapes (which are less fragmented), so the left-skew and increasing linear trend of the fitted model may be due to the disproportionate proportion of fragmented forest near villages relative to primary forest. In Fig. 35, canopy tree species diversity is highest at edge density values around 200-250, which are intermediate to the full range (0-500). We speculate that this could occur because niche partitioning encourages recruitment of tree species that are different than those that dominate primary forest patches, and

also different than those that dominate extensive clearings, both of which have low edge density. On the upper end of the disturbance scale, patches with higher fragmentation levels may have lower diversity values because of insufficient niche differentiation. Fig. 36 suggests canopy tree species diversity is maximized at a log splitting index between 1 and 2 (which are intermediate to the full range, 0-3) where patches of tree species are sufficiently fragmented for the recruitment of additional, and diverse species but also sufficiently partitioned. However, the inflection of the convex curve here is not extremely negative. As mentioned in section 4.2, splitting index is a useful metric to include due to its insensitivity to the inclusion or omission of small patches, and given the number of artificially small clusters generated by the automated classification is an artifact of processing, this property gives splitting index an advantage over other common measures of subdivision like patch density and mean patch size.

The models presented here resulted from preliminary attempts at modeling the spectral species and fragmentation statistics. In general, for all the models, further testing and optimization of the spectral species identification parameters and use of the supervised classification may improve our ability to understand whether the relationships support the intermediate disturbance hypothesis. Several improvements on the models are also possible. For example, the outcome variable of the simulation (Figure 1) is a count of the number of resources that are available, and outcome variable in the fitted models (Figures 32-36) is the (exponentiated) Shannon diversity of the spectral species, which is a proxy for the number of equally common species or "effective number of species" (Jost 2006). In both cases, it may be better to use a different distribution to model the responses; the Poisson distribution for the simulated data, and the gamma, log-normal, or exponential distribution to model Shannon diversity. In addition, we could use informed priors and multilevel models (hyperpriors) to further improve the modeling results. Overall, however, our preliminary results provide tentative quantitative support for the claim that Q'eqchi' forest clearing and swidden cultivation enhances forest biodiversity in Southern Belize, especially at large spatial scales of analysis.

6. Conclusions

The results of this project could help to reveal the patterns and underlying mechanisms that drive environmental outcomes of subsistence practices. Improving existing models of subsistence

agriculture could have significant impacts on policy directed at people and working forests worldwide. While our work on this project has not yet reached its end, there is much to appreciate in the progress we have made as a team. Beyond quantitative results, this thesis should serve as a blueprint for collaboration and continued learning across domains of knowledge. The scale and scope of this project is beyond what could have been reasonably conceived and carried out by one researcher; the process of meticulous, thoughtful, interdisciplinary research necessitates creating communities of shared expertise and constant learning. Navigating the intricacies and nuances of work that was not my own was a challenge, but it did not take long for me to start owning parts of the project, which taught me the importance of thorough research and presenting findings in a way that creates generative, inclusive conversations about the trajectory of future work.

In adapting data science methodologies to anthropological questions, I am lucky to have not only broadened my technical skill set, but acquired an appreciation for not knowing all the answers. As one part of a larger project, the creation of this thesis allowed me to reflect on what it means to be a data scientist and a researcher; the importance of listening, failures and success, and thinking critically about how your work can benefit the people we are privileged to study as anthropologists. The models we rely on may be just as human as those who created them and gave them life, filled with biases and flaws. But maybe also empathy, if we can look beyond the ivory towers and create from a place of community and healing rather than gain. In the future, we will all undoubtedly be data scientists, at least to some degree, as a consequence of the increasingly data-driven world we live in. First and foremost, however, we are all storytellers and artists, and always have been; we should not forget this. When we bring our full identities to the work we perform, we do research not *on* others, but *for* them.

Acknowledgements

I would like to thank Dr. Sean Downey, Shane Scaggs, and all involved in HCL for their guidance and investment in my skills over the last year, as well as Dr. Bill Peterman and Dr. Mark Moritz for serving on the committee for my Honors thesis defense. We are especially indebted to the drone pilot Jacob Moschler as well as local experts of Crique Sarco and Graham Creek for sharing their knowledge and assisting with data collection. We thank Dr. Rongjun Qin (The Ohio State University) and Shuang Song (The Ohio State University) for curating the photogrammetry stitching process and lending their expertise in remote sensing. The study was funded by the National Science Foundation (NSF#1818597, NSF#1553875). This study was done under permission and current laws of the government of Belize.

References

Balée, W. (2013). Cultural forests of the Amazon: a historical ecology of people and their landscapes. University of Alabama Press.

Berkes, Fikret. Sacred Ecology. Routledge, 2017.

Bird, R. Bliege, Bird, D. W., Codding, B. F., Parker, C. H., Jones, J. H. (2008). The "fire stick farming" hypothesis: Australian Aboriginal foraging strategies, biodiversity, and anthropogenic fire mosaics. Proceedings of the National Academy of Sciences Sep 2008, 105 (39) 14796-14801; DOI: 10.1073/pnas.0804757105

Bird, Rebecca Bliege. "Disturbance, complexity, scale: New approaches to the study of human–environment interactions." Annual Review of Anthropology 44 (2015): 241-257.

Boserup, E. (1965). The conditions of agricultural growth: the economics of agrarian change under population pressure. G. Allen and Unwin.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5-32.

Bürkner P (2017). "brms: An R Package for Bayesian Multilevel Models Using Stan." Journal of Statistical Software, 80(1), 1–28. doi: 10.18637/jss.v080.i01.

Codding, Brian & Bliege Bird, Rebecca & Kauhanen, Peter & Bird, Douglas. (2014). Conservation or Co-evolution? Intermediate Levels of Aboriginal Burning and Hunting Have Positive Effects on Kangaroo Populations in Western Australia. Human Ecology. 42. 10.1007/s10745-014-9682-4.

Connell, Joseph H. "Diversity in Tropical Rain Forests and Coral Reefs." *Science*, vol. 199, no. 4335, 1978, pp. 1302–1310.

Conklin, H. C. (1954). Section of anthropology: an ethnoecological approach to shifting agriculture. Transactions of the New York Academy of Sciences, 17(2 Series II):133–142.

Cox, K., Gerwing, T.G., Macdonald, T., Hessing‐Lewis, M., Millard-Martin, B., Command, R.J., Juanes, F., & Dudas, S.E. (2019). Infaunal community responses to ancient clam gardens. Ices Journal of Marine Science, 76, 2362-2373.

Downey, S. S. (2015). Q'eqchi' Maya Swidden Agriculture, Settlement History, and Colonial Enterprise in Modern Belize. Ethnohistory, 62(4), 751-779.

Downey, Sean S. "Can properties of labor-exchange networks explain the resilience of swidden agriculture?." Ecology and Society 15.4 (2010).

Downey, Sean S., Drew Gerkey, and Shane A. Scaggs. "The Milpa Game: a Field Experiment Investigating the Social and Ecological Dynamics of Q'eqchi' Maya Swidden Agriculture." Human Ecology 48.4 (2020): 423-438.

Féret, J.-B., Asner, G.P., 2014. Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. Ecol. Appl. 24, 1289–1296. https://doi.org/10.1890/13-1824.1

Féret, J.-B., de Boissieu, F., 2019. biodivMapR: an R package for α‐ and β‐diversity mapping using remotely‐sensed images. Methods Ecol. Evol. 00:1-7. https://doi.org/10.1111/2041-210X.13310

Food and Agriculture Organization. (2010) Global Forest Resources Assessment 2010 main report, FAO For. Pap. 163, 378 pp., Rome. http://www.fao.org/forestry/fra/fra2010/en/

Forman, R. T. T., and M. Godron. 1986. Landscape Ecology. John Wiley & Sons, New York.619 pp.

Fox, Jeremy W. "The intermediate disturbance hypothesis should be abandoned." Trends in ecology & evolution 28.2 (2013): 86-92.

Geertz, Clifford. Agricultural involution: The processes of ecological change in Indonesia. Vol. 11. Univ of California Press, 1963.

Hesselbarth, M. H. K., Sciaini, M., With, K. A., Wiegand, K., Nowosad, J. 2019. landscapemetrics: an open‐source R tool to calculate landscape metrics. – Ecography 42:000–000 (ver. 1.5.1).

Jaeger, Jochen. (2000). Landscape division, splitting index, and effective mesh size: New measures of landscape fragmentation. Landscape Ecology. 15. 115-130. 10.1023/A:1008129329289.

Jost, Lou. "Entropy and diversity." Oikos 113.2 (2006): 363-375.

Kindlmann, P., Burel, F. Connectivity measures: a review. *Landscape Ecol* **23,** 879–890 (2008). https://doi.org/10.1007/s10980-008-9245-4

Levin, S.A. (1992), The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. Ecology, 73: 1943-1967. https://doi.org/10.2307/1941447

McElreath, R., O'Reilly for Higher Education (Firm),, & Safari, an O'Reilly Media Company. (2020). *Statistical Rethinking, 2nd Edition.*

McGarigal, Kevin. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. Vol. 351. US Department of Agriculture, Forest Service, Pacific Northwest Research Station, 1995.

Micasense corp. Micasense RedEdge 3 Multispectral Camera User Manual. Rev 06–October 2015. https://micasense.com/.

Moschler, Jacob D. Distributed Search Method for Teams of Small Unmanned Aircraft Systems. Diss. 2018.

Ohio Supercomputer Center. 1987. Ohio Supercomputer Center. Columbus OH: Ohio Supercomputer Center. http://osc.edu/ark:/19495/f5s1ph73.

Oldeland, J., Wesuls, D., Rocchini, D., Schmidt, M., & Jürgens, N. (2010). Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity?. Ecological Indicators, 10(2), 390–396. https://doi.org/10.1016/j.ecolind.2009.07.012

Osman, R. W. (2008).The Intermediate Disturbance Hypothesis. Encyclopedia of Ecology (Second Edition), 441-450. https://doi.org/10.1016/B978-0-12-409548-9.09480-X

Patton, D. R. 1975. A diversity index for quantifying habitat "edge". Wildl. Soc. Bull. 3:171-173.

Robbins, C. S., D. K. Dawson, and B. A. Dowell. 1989. Habitat area requirements of breeding forestbirds of the middle Atlantic states. Wildl. Monogr. 103. 34 pp.

Rocchini, Duccio. (2007). Effects of spatial and spectral resolution in estimating ecosystem α-diversity by satellite imagery. Remote Sensing of Environment. 111. 423-434. 10.1016/j.rse.2007.03.018.

Roxburgh, S. H., Shea, K., & Wilson, J. B. (2004). The intermediate disturbance hypothesis: patch dynamics and mechanisms of species coexistence. Ecology, 85(2), 359-371.

Saavedra, F., Hensen, I., Beck, S. G., Böhning-Gaese, K., Lippok, D., Töpfer, T., & Schleuning, M. (2014). Functional importance of avian seed dispersers changes in response to human-induced forest edges in tropical seed-dispersal networks. Oecologia, 176(3), 837-848.

Taylor, P., Fahrig, L., Henein, K., & Merriam, G. (1993). Connectivity Is a Vital Element of Landscape Structure. Oikos, 68(3), 571-573. doi:10.2307/3544927

Tischendorf, L. and Fahrig, L. (2000), On the usage and measurement of landscape connectivity. Oikos, 90: 7-19. https://doi.org/10.1034/j.1600-0706.2000.900102.x

Weinstock, J. A. (2015). The future of swidden cultivation. In Cairns, M. F., (ed.), Shifting cultivation and environmental change: indigenous people, agriculture, and forest conservation. Routledge, pp. 179–85.

Wilkinson, D. (1999). The Disturbing History of Intermediate Disturbance. *Oikos, 84*(1), 145-147. doi:10.2307/3546874

Appendix

1. Ground truthed biodiversity transects (data provided by Shane S. Scaggs):

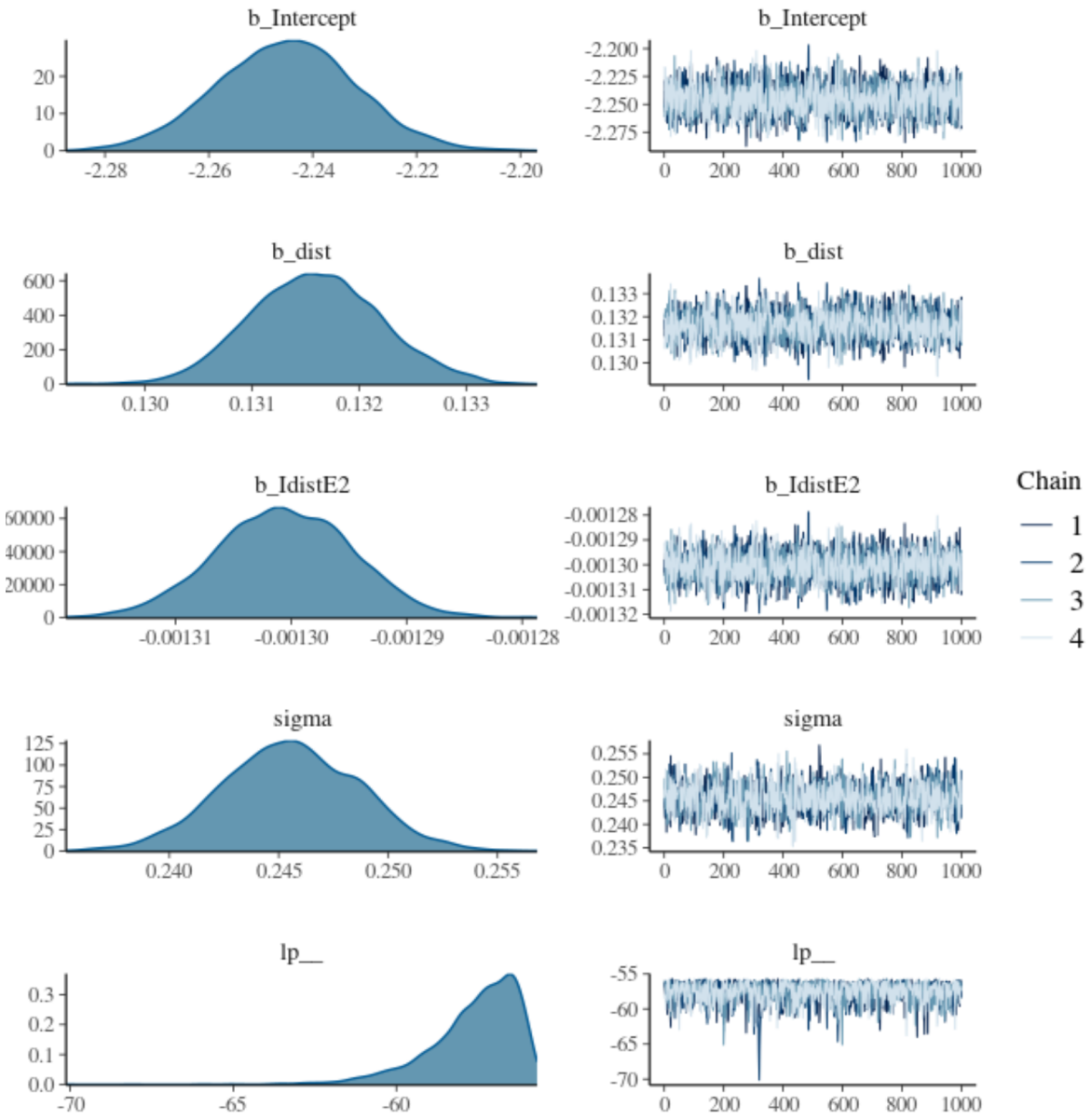
Class	Num species counted	Species Richness
huamil 20-year	354	46
huamil 4-year	287	41
huamil 8-year	413	50
primary (>20)	793	98

2. Convergence plots for Bayesian models

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: res.s ~ 1 + dist + I(dist^2)
   Data: d (Number of observations: 3000)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept    -2.25      0.01    -2.27    -2.22 1.00     2029     2391
dist          0.13      0.00     0.13     0.13 1.00     2425     2014
IdistE2      -0.00      0.00    -0.00    -0.00 1.00     2471     2251

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.25      0.00     0.24     0.25 1.00     1663     1405
```



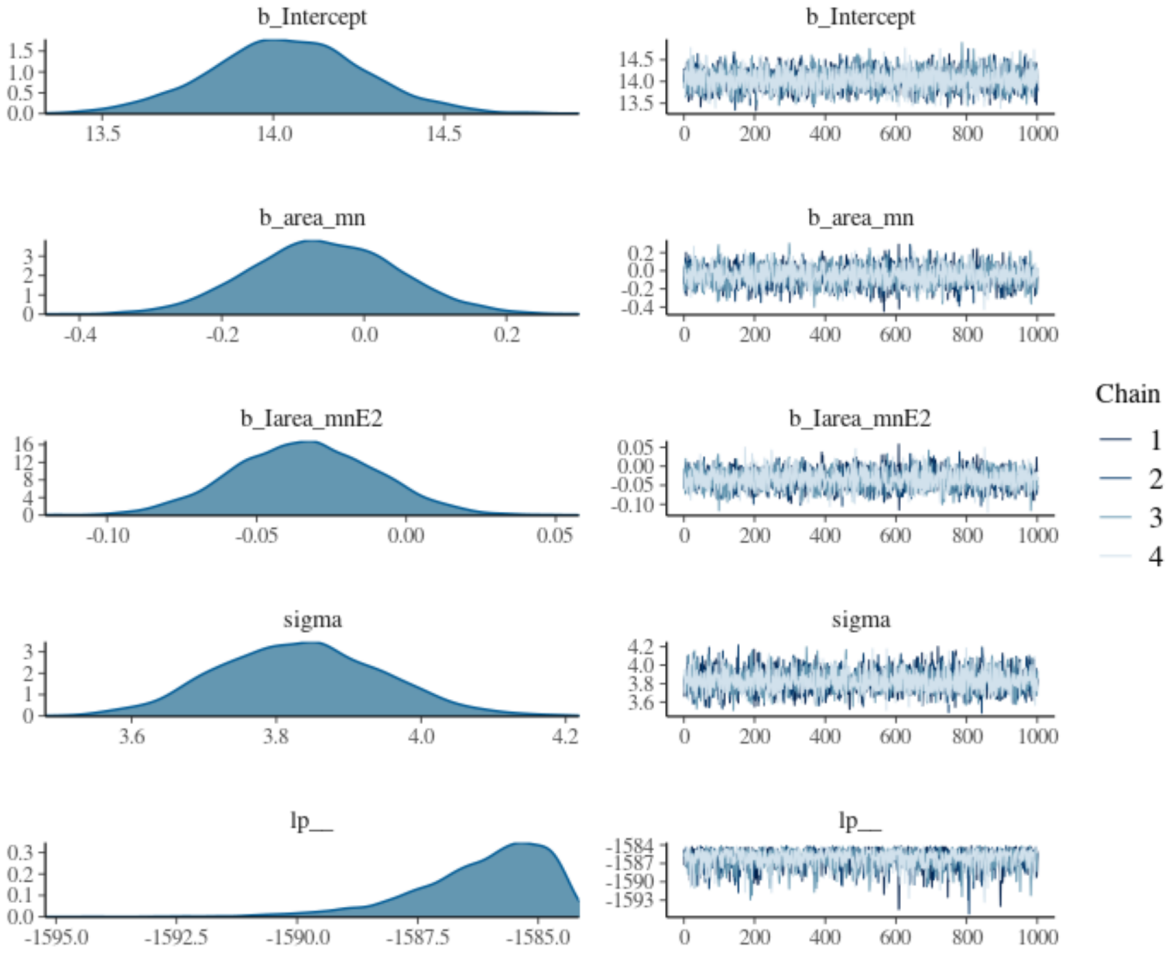
2.1. RESOURCES~DISTURBANCE (simulation)

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: ss.shdi ~ 1 + area_mn + I(area_mn^2)
   Data: m (Number of observations: 569)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     14.04      0.22    13.60    14.49 1.00     4606     3255
area_mn       -0.06      0.10    -0.26     0.15 1.00     3416     2808
Iarea_mnE2    -0.03      0.02    -0.08     0.01 1.00     3495     3128

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     3.83      0.11     3.62     4.06 1.00     3991     2697

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

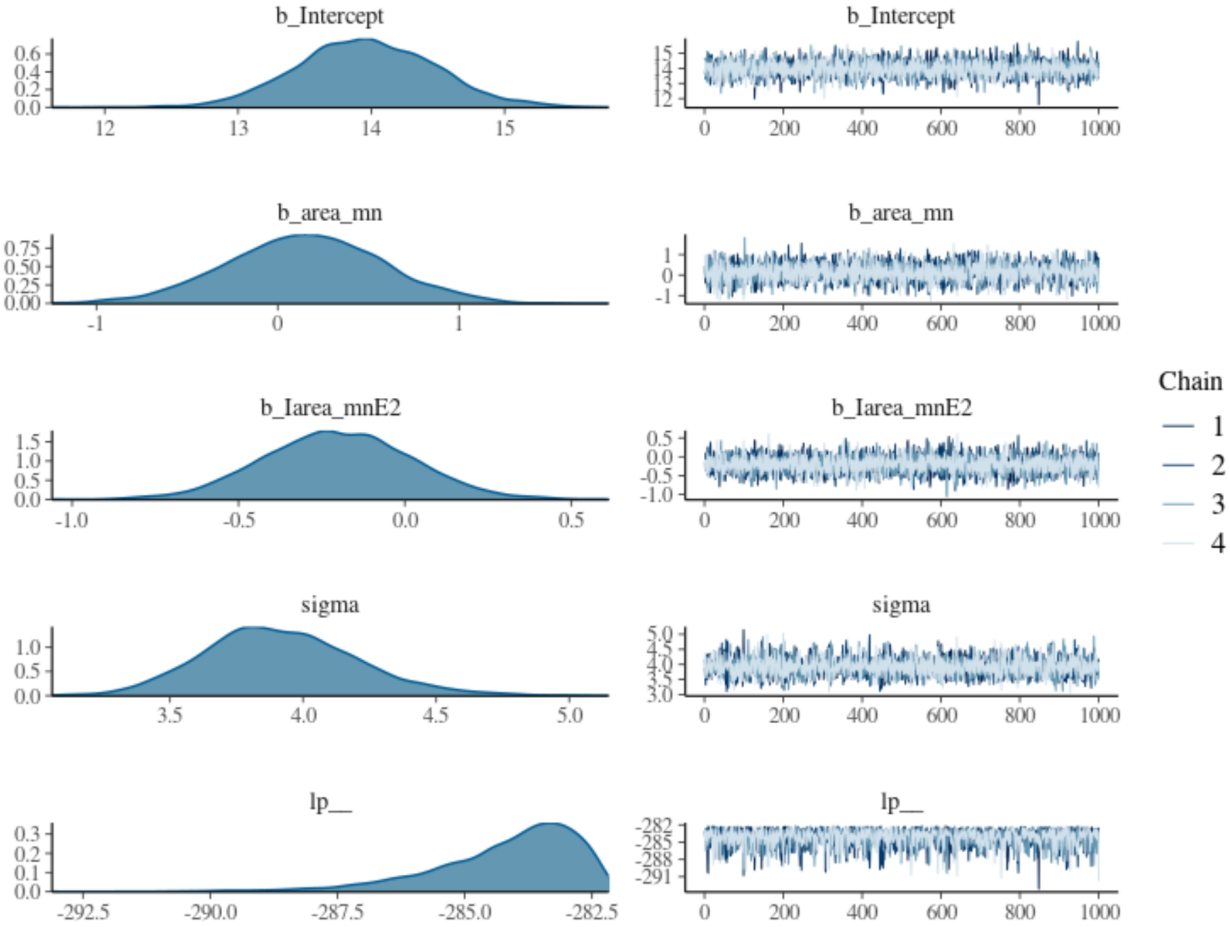


2.2. SHDI~AREA_MN (class-level)

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: ss.shdi ~ 1 + area_mn + I(area_mn^2)
   Data: m (Number of observations: 97)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     13.96      0.52    12.97    15.00 1.00     3994     3079
area_mn        0.15      0.42    -0.69     0.98 1.00     4138     3111
Iarea_mnE2    -0.20      0.22    -0.64     0.23 1.00     4131     3093

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     3.92      0.29     3.40     4.53 1.00     4149     2838
```

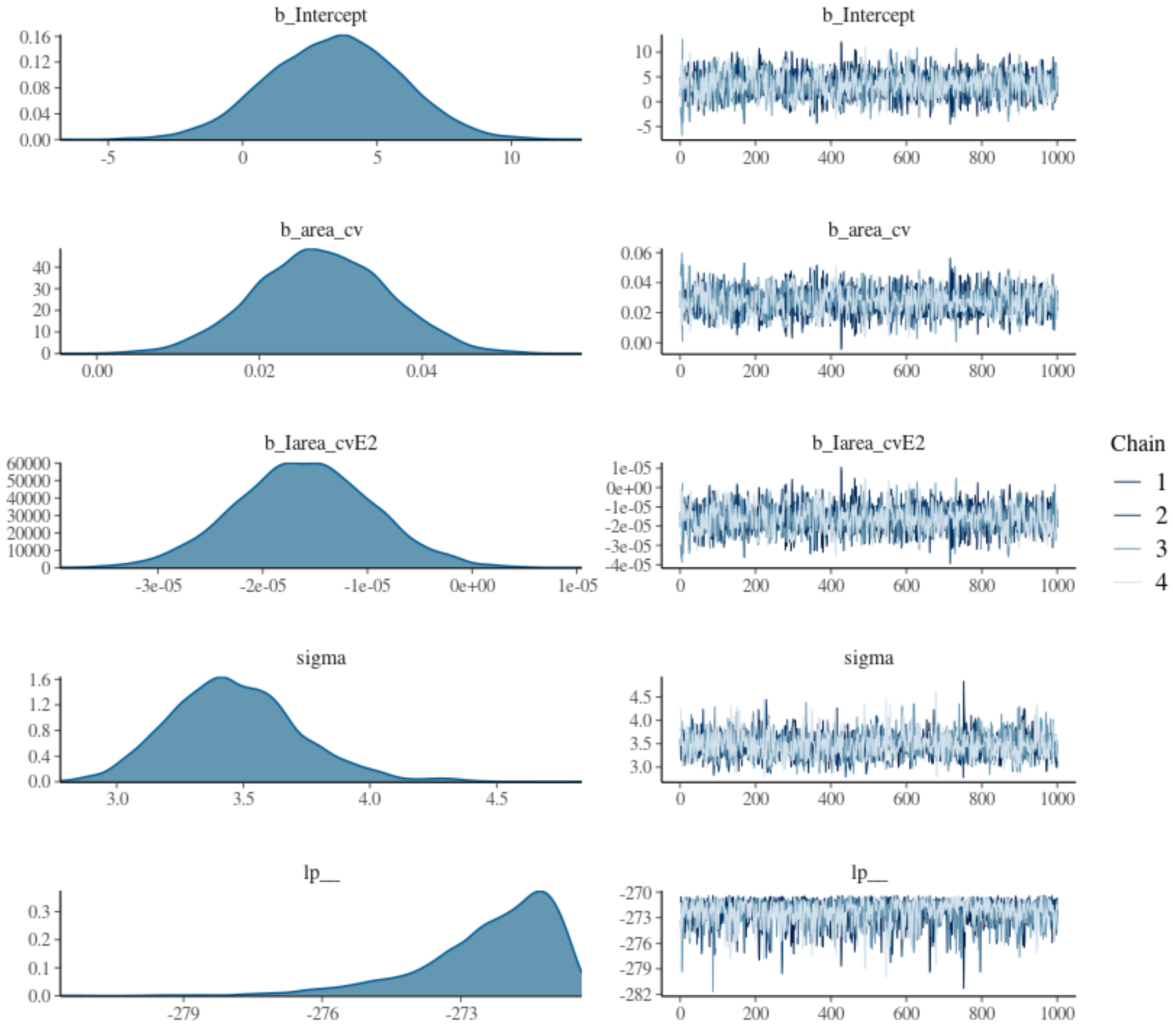


2.3. SHDI~AREA_MN (landscape-level)

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: ss.shdi ~ 1 + area_cv + I(area_cv^2)
   Data: m (Number of observations: 97)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept      3.42      2.45    -1.37     8.12 1.00     1877     1887
area_cv        0.03      0.01     0.01     0.04 1.00     1841     1582
Iarea_cvE2    -0.00      0.00    -0.00    -0.00 1.00     1943     1810

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     3.47      0.26     3.02     4.02 1.01     1077     1067
```

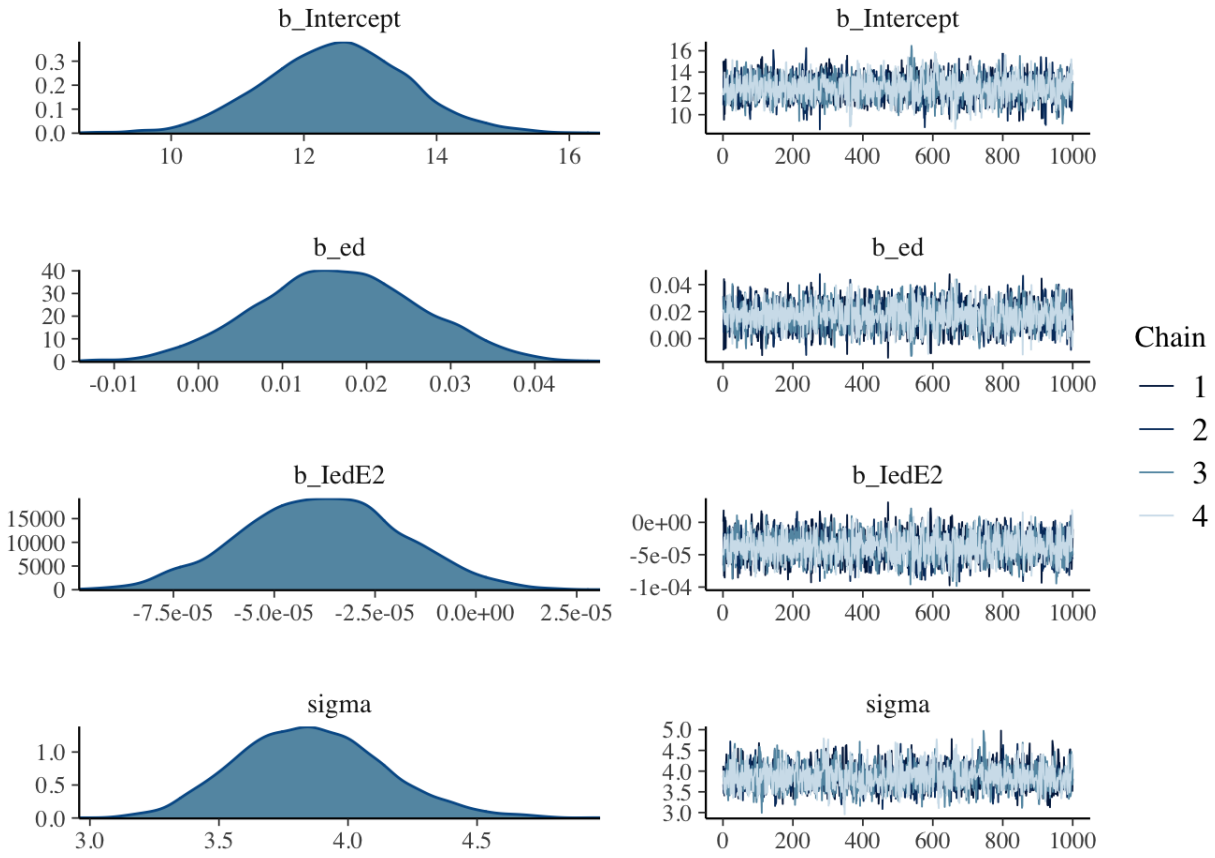


2.4. SHDI~AREA_CV

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: ss.shdi ~ 1 + ed + I(ed^2)
   Data: m (Number of observations: 97)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept    12.49      1.07    10.40    14.63 1.00     2257     2054
ed            0.02      0.01    -0.00     0.04 1.00     2296     2080
IedE2        -0.00      0.00    -0.00     0.00 1.00     2432     2281

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     3.86      0.28     3.35     4.45 1.00     2207     1860
```

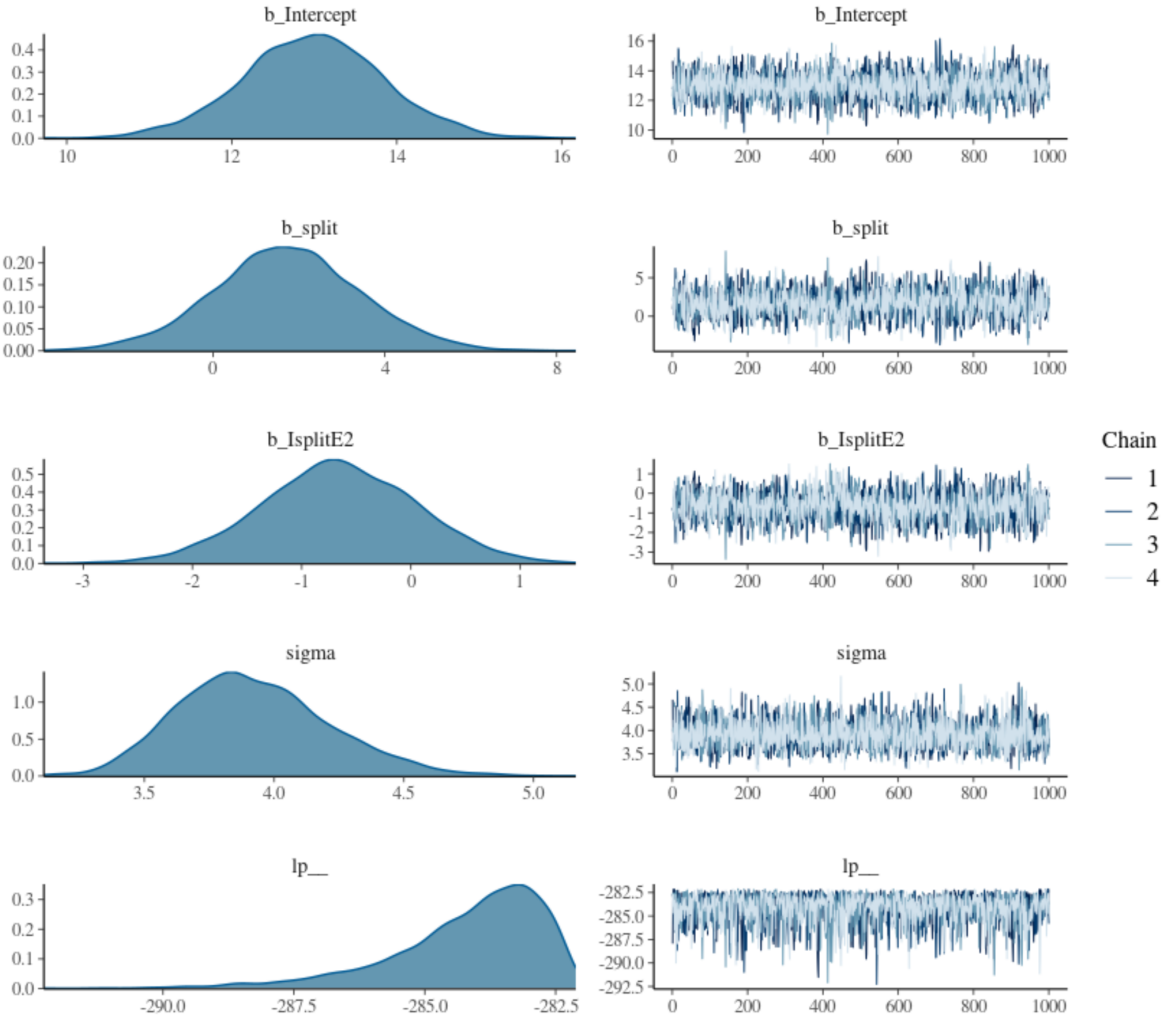


2.5. SHDI~ED

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: ss.shdi ~ 1 + split + I(split^2)
   Data: m (Number of observations: 97)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept    12.98      0.86    11.23    14.70 1.00     2544     2747
split         1.72      1.72    -1.73     5.13 1.00     1866     1842
IsplitE2     -0.66      0.71    -2.06     0.71 1.00     1840     2051

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     3.92      0.29     3.42     4.54 1.00     2725     2579
```



2.6. SHDI~SPLIT